Bayesian Inference for High-Dimensional Data with Applications to Portfolio Theory

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach Mathematik

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

der Humboldt-Universität Berlin

von

David Werner Bauder

Präsidentin der Humboldt-Universität Berlin:

Prof. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:

Prof. Dr. Elmar Kulke

Gutachter

1. Gutachter: PD Dr. habil. Taras Bodnar
2. Gutachter: Prof. Dr. Wolfgang Schmid
3. Gutachter: Prof. Dr. Joanna Tyrcha

Datum der mündlichen Prüfung: 26.11.2018

Zusammenfassung

Die Gewichte eines Portfolios liegen meist als als Kombination des Produkts der Präzisionsmatrix und des Erwartungswertvektors vor. In der Praxis müssen diese Parameter geschätzt werden, allerdings ist die Beschreibung der damit verbundenen Schätzunsicher-heit über eine Verteilung dieses Produktes eine Herausforderung. In dieser Arbeit wird demonstriert, dass ein geeignetes bayesianisches Modell nicht nur zu einer leicht zugäng-lichen Posteriori-Verteilung führt, sondern auch zu leicht interpretierbaren Beschreibungen des Portfoliorisikos, wie beispielsweise einer Ausfallwahrscheinlichkeit des gesamten Portfolios zu jedem Zeitpunkt.

Dazu werden die Parameter mit ihren konjugierten Prioris ausgestatet. Mit Hilfe bekannter Ergebnisse aus der Theorie multivariater Verteilungen ist es möglich, eine stochastische Darstellung für relevante Ausdrücke wie den Portfoliogewichten oder des effizienten Randes zu geben. Diese Darstellungen ermöglichen nicht nur die Bestimmung von Bayes-Schätzern der Parameter, sondern sind auch noch rechentechnisch hoch effizient, da Zufallszahlen nur aus bekannten und leicht zugänglichen Verteilungen gezogen werden. Insbesondere aber werden Markov-Chain-Monte-Carlo Methoden nicht benötigt.

Angewendet wird diese Methodik an einem mehrperiodigen Portfoliomodell für eine exponentielle Nutzenfunktion, am Tangentialportfolio, zur Schätzung des effizienten Randes, des globalen Minimum-Varianz-Portfolios wie auch am gesamten Mittelwert-Varianz Ansatzes. Für alle behandelten Portfoliomodelle werden für wichtige Größen stochastische Darstellungen oder Bayes-Schätzer gefunden. Die Praktikabilität und Flexibilität wie auch bestimmte Eigenschaften werden in Anwendungen mit realen Datensätzen oder Simulationen illustriert.

Abstract

Usually, the weights of portfolio assets are expressed as a comination of the product of the precision matrix and the mean vector. These parameters have to be estimated in practical applications. But it is a challenge to describe the associated estimation risk of this product. It is demonstrated in this thesis, that a suitable Bayesian approach does not only lead to an easily accessible posteriori distribution, but also lead to easily interpretable risk measures. This also includes for example the default probability of the portfolio at all relevant points in time.

To approach this task, the parameters are endowed with their conjugate priors. Using results from the theory of multivariate distributions, stochastic representations for the portfolio parameter are derived, for example for the portfolio weights or the efficient frontier. These representations not only allow to derive Bayes estimates of these parameters, but are computationally highly efficient since all th necessary random variables are drawn from well known and easily accessible distributions. Most importantly, Markov-Chain-Monte-Carlo methods are not necessary.

These methods are applied to a multi-period portfolio for an exponential utility function, to the tangent portfolio, to estimate the efficient frontier and also to a general mean-variance approach. Stochastic representations and Bayes estimates are derived for all relevant parameters. The practicability and flexibility as well as specific properties are demonstrated using either real data or simulations.

"Woran arbeiten Sie?" wurde Herr K. gefragt. Herr K. antwortete:

"Ich habe viel Mühe, ich bereite meinen nächsten Irrtum vor."

Bertold Brecht, Geschichten vom Herrn Keuner.

Acknowledgements

It is impossible to thank everyone in an appropriate manner to whom I am grateful. I would like to thank Taras Bodnar for his supervision and support in particular. There was never a meeting after which I felt under pressure or bad - but always motivated and eager to try out new ideas. Pursuing a PhD under your guidance was a pleasure. Thank you for always having an open ear.

Furthermore, I'd like to thank all my coauthors Rostyslav Bodnar, Stepan Mazur, Yarema Okhrin, Nestor Parolya and Wolfgang Schmid. Financial support by the German Science Foundation (DFG) via the projects BO 3521/3-1 and SCHM 859/13-1 is thankfully acknowledged.

My deepest gratitute goes to my parents Anita and Hartmut Bauder for trusting and supporting me with all my studies, endeavours, adventures and decisions. This includes everything not always related to my education. Most importantly for being not only parents but very close friends. Of course, this includes my sister Mira. I owe a great part of my confidence and my positive attitude towards life to all of you.

I also have to thank very dear friends for all the discussions. Their influence during all my studies can not be neglected. This includes particularly (in alphabetical order) Dominik, Hendrik, Jana, Marco and Wolfgang. Thank you for the challenging, but warm and supporting environment.

My girlfriend and partner Christiane is a huge source of tranquility, relaxation and support. Thank you for having no clue what my thesis is about.

Contents

List of Figures

Chapter 1

Preface

Methods for dealing with the phenomenon of Big Data gain lots of attention in a variety of fields related to data analysis. Finding methods with a high prediction potency seem to be desirable to develop. Statistical uncertainty does not seem to be as popular, although it is of high relevance: when someone makes a prediction, the accuracy of this prediction is also of interest. For example, think of a car with a very sophisticated navigation method where it is possible that the driver gives the car complete driving autonomy: if the car gets in a traffic situation which is not covered by a sufficient amount of learning data, the prediction can not be accurate enough and the car should give back the control to the driver. Clearly, it is necessary to quantify this complex form of statistical uncertainty.

We do not consider cars or driving systems in this thesis, but specific portfolio-models. A portfolio usually involes an investment decision. Since an investment can be regarded as a quite committed form of prediction, one might naturally be interested in quantifying the related portfolio risk. This portfolio risk is usually restricted to economic risk about the future behaviour of asset returns. The extensive losses of nearly all stock markets in the world during the financial crisis of 2008 is a good illustration for this but the vast returns after the recovery of the markets illustrates this type of risk as well. Unfortunately, this is not an adequate description of a portfolio's risk in practice. Since the parameters of a portfolio are unknown in practice, they have to be estimated. Hence, to fully describe a portfolio's risk in practice, it is of paramount importance to account for risk resulting from estimating the parameters.

This estimation risk leads to suboptimal portfolio choices. Usually, applying the mean-variance paradigm introduced by Markowitz (1952) involves two steps: in the first step, the parameters are estimated. The second step is solving the portfolio problem, treating the estimations as true parameters. Of course, there exist a vast amount of portfolio models and extensions since Markowitz (1952), but this two-step-approach appears to be accepted in practice as well as in research. Although this procedure seems quite simple, there are considerable problems or even obstacles in practice as for example described by Hodges and Brealey (1978), Michaud (1989), Best and Grauer (1991), Barberis (2000) and Pástor (2000). Since the 70's, this two-step approach is critisized, for example in Barry (1974), Brown (1976), Klein and Bawa (1976) or Jobson and Korkie (1980) and, for a more general and more modern overview, in Best and Grauer (1991) and Litterman (2003). There is evidence that the estimation risk of the parameters can not be neglected: Britten-Jones (1999) demonstrated that the sampling error of important portfolio parameters can be exceedingly large. Similarly, results in Gibbons et al. (1989), Shanken (1992), Okhrin and Schmid (2006), Bodnar and Schmid (2008a), Bodnar and Schmid (2009) and Bodnar and Schmid (2011) point in the same direction for different portfolio models. To deal with this issue, it might be of interest whether or not an investment in an asset might be significant as in French and Poterba (1991) or if an investment deviates from a prespecified value as in Britten-Jones (1999) or Ang and Bekaert (2002). It is also shown how to test the sensitivity of the asset weighting to changes in the underlying parameters as in Chopra and Ziemba (1993) or Bodnar (2009).

One of the first methods to approach estimation risk were proposed by Winkler (1973), Barry (1974), Winkler and Barry (1975) and Bawa et al. (1979) who followed a Bayesian approach by applying a non-informative prior to the parameters or used a predictive distribution to track the estimation risk. See Bawa et al. (1979) for a review on early examples where Bayesian methods are applied in portfolio theory. Jobson and Korkie (1980), Jorion (1985), Jorion (1986) and Frost and Savarino (1986) used empirical Bayes estimates to shrink estimated parameters to a specified values. Wang (2005), Kan and Zhou (2007), Golosnoy and Okhrin (2007), Golosnoy and Okhrin (2008) and Bodnar et al. (2017c) took a similar shrinkage-approach. Hence, the Bayesian approach can be regarded as an established method in portfolio analysis. As Avramov and Zhou (2010) point out, Bayesian methodology resembles human decision making - updating prior beliefs by experience or data, respectively. Since investing is still a decision made by humans, at least up to a certain degree, this standpoint might be compelling. But speaking in more practical terms, the distribution of a random variable does not require asymptotic arguments when only finite samples are available. But besides that Bayesian statistics account properly for parameter and model uncertainty in a practical way, this method has a deep philosophical and mathematical foundation. This is briefly discussed in the second section of this preface. Besides this, we want to focus on practical advantages which stem directly from the distribution of the unknown parameters when data is available, the posterior distribution.

Unfortunately, it can be quite challenging to access this posterior distribution. In many

The literature is discussed in more detail in the introductory sections of the respective papers of this thesis.

cases, Markov Chain Monte Carlo (MCMC) methods are used simulate the posterior distribution. While MCMC-procedures make the posterior distribution accessible, they often require huge amounts of computational ressources. We do not make use of MCMC-methods once in all the papers associated with this thesis. Instead we access the posterior distribution of all the parameters using a stochastic representation. Stochastic representations are a well-established technique to describe the distributions of multi- or matrixvariate random variables, see Muirhead (1982) or Gupta et al. (2013). Bodnar and Schmid (2011) already used a stochastic representation to derive the distribution of important portfolio parameters in a frequentist setting.

We endow the mean and the covariance matrix with their conjugate priors and derive the stochastic representations of the parameters of four important portfolio models: a multi-period portfolio for the exponential utility function as solved in Bodnar et al. (2015a), the Global Minimum Variance portfolio, the efficient frontier and to the mean-variance paradigm in general introduced by Markowitz (1952). To prepare all the following results, the next section presents a short introduction to the Bayesian approach to statistics. After that, an intuitive recapitulation of portfolio theory is given. The last section gives an outline to the thesis and reviews the contributions made in the underlying research papers.

1.1 The Bayesian approach to statistics

In this section we provide a brief construction and motivation of the Bayesian approach and also present the concepts which reappear in this thesis. One of the main differences between frequentist and Bayesian statistics is the assumption of independent observations. If we want to use aggregated data to make predictions about future observations and assume independent observations, then we would treat every new data point individually. Past observations are not relevant since

$$p(y_1, ..., y_n) = \prod_{i=1}^{n} p(y_i) \tag{1.1}$$

actually implies that the probability for future observations conditional on past observations does not depend on past observations, hence

$$p(y_{n+1}, ..., y_m | y_1, ..., y_n) = p(y_{n+1}, ..., y_m) \tag{1.2}$$

While this assumption is practical, it does not describe data very well. A slightly weaker assumption is the assumption of *exchangeability*.

Definition 1. *A sequence of random variables $y_1, y_2, ..., y_n$ is finitely exchangeable if*

$$y_1, y_2, ..., y_n \stackrel{d}{=} y_{\pi(1)}, y_{\pi(2)}, ..., y_{\pi(n)} \tag{1.3}$$

for every permutation π on $\{1, ..., n\}$. An infinite sequence of random variables $y_1, y_2, ...$ is said to be infinitely exchangeable if every finite subsequence is finitely exchangeably.

The symbol $\stackrel{d}{=}$ describes equality in distribution. This definition of exchangeability also implies, that a sequence of independent random variables is also exchangeable, but exchangeability does not imply independence. The intuition behind this theorem is that the order in which we encounter data is not of interest or relevance, leading to simplificated inference procedures. Nevertheless, exchangeability of course is sometimes also too restrictive. For such sequences which can not be considered to be exchangeable, it is possible to use auxiliary information to partition the sequences into exchangeable sets. For example, consider two dice. The first one is fair and used in a casino on weekends and the second one is biased and used during the week. Then the data of all throws is exchangeable within the set of throws during the week and on weekends. But of course, there exist many extensions and variations to Definition 1 which all aim at grouping the data into exchangeable sets for easier inference.

A very prominent and important result from the assumption of exchangeability is the idea that an infinite sequence of random variables $y_1, y_2, ...$ is exchangeable if and only if there exists a random probability measure ν with respect to which the considered sequence of random variables $y_1, y_2, ...$ is conditionally independently and identically distributed with their distribution being the random probability measure ν. This finding is most prominently known as *de Finetti's theorem.*

Theorem 1. *Let $y_1, y_2, ...$ be an infinitely exchangeable sequence of binary random variables with probability measure P. Then there exists a distribution function Q such that the joint probability mass function $p(y_1, ..., y_n)$ defined by the measure P is given as*

$$p(y_1, .., y_n) = \int_{-\infty}^{\infty} \prod_{i=1}^{n} \varphi^{y_i} (1 - \varphi)^{1-y_1} dQ(\varphi). \tag{1.4}$$

where Q is the distribution function of the limiting empirical frequency

$$\theta \stackrel{a.s.}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1.5}$$

and $\theta \sim Q$.

De Finetti proved this finding in the case of binary random variables in deFinetti (1931) and was extended especially by Hewitt and Savage (1955) or Ryll-Nardzewski (1957). A proof in more modern terms is presented in Bernardo and Smith (2000). In this original form, De Finetti's theorem can be interpreted as regarding the elements of the binary exchangeable sequence as independent realizations of a Bernoulli-distribution with success probability θ, where $\theta \sim Q$. This distribution Q can be regarded as the Belief about the limiting empirical frequency of successes in the data. From the Bayesian perspective, Q can be seen as a motivation for a prior distribution. The general form of de Finetti's theorem was derived by Hewitt and Savage (1955):

Theorem 2. *Let $y_1, y_2, ...$ be an exchangeable sequence of real-valued random variables with probability meausure P. Then there exists a probability measure μ on the set of probability measures $\mathcal{P}(\mathbb{R})$ on $\mathbb{R}$, such that*

$$p(y_1 \in A_1, ..., y_n \in A_n) = \int_{\mathcal{P}(\mathbb{R})} \prod_{i=1}^{n} \Phi(A_i)\mu(d\Phi). \tag{1.6}$$

It further holds that μ is the distribution function of a probability measure ν, where ν is defined by the limiting empirical measure:

$$\nu(B) \overset{a.s.}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_B(y_i) \tag{1.7}$$

where $\nu \sim \mu$ and B covers all elements of the Borel σ-algebra and $\mathbb{I}$ is the indicator function.

This general form of de Finetti's theorem shows that if $y_1, y_2, ...$ are infinitely exchangeable, then there exists a measure μ on measures in a way that $\nu \sim \mu$ and that $y_i \mid \nu \overset{iid}{\sim} \nu$. While the original form of de Finetti's theorem in Theorem 1 states that the existing random probability measure ν on $\{0, 1\}$ can uniquely be described by the parameter θ, the extention by Hewitt and Savage (1955) in Theorem 2 introduces a random measure concentrated on $\{1, ..., k\}$ and is uniquely defined by a $(k-1)$-dimensional parameter. Although it would be possible to consider an arbitrarily complicated random probability measure ν, even up to infinitely many parameters in case of the Dirichlet process, the cases considered in this thesis are finite. If ν is almost surely a multivariate normal distribution, then ν is fully characterized by the mean and the covariance matrix. In fact, in case of a finite dimensional ν, there exists a distribution function Q such that the joint density of $y_1, ..., y_n$ is given as

$$p(y_1, ..., y_n) = \int_{\Theta} \prod_{i=1}^{n} p(y_i|\varphi)dQ(\varphi), \tag{1.8}$$

where $p(\cdot|\varphi)$ is the density corresponding to $\varphi \in \Theta$. Furthermore, we can formulate a predictive densitiy, given by

$$p(y_{m+1},...,y_n|y_1,...,y_m) = \int_\Theta \prod_{i=m+1}^{n} p(y_i|\varphi)dQ(\varphi|y_1,...y_m), \tag{1.9}$$

with

$$dQ(\theta|y_1,...,y_m) = \frac{\prod_{i=1}^m p(y_i|\theta)dQ(\theta)}{\int_\Theta \prod_{i=1}^m p(y_i|\varphi)dQ(\varphi)} \tag{1.10}$$

It is noteworthy that the view of the underlying random parameter θ yielding i.i.d-data is still the same. The prior belief $Q(\theta)$ was updated, yielding a posterior belief $Q(\theta|y_1,...,y_m)$ using *Bayes rule*,

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_\Theta p(y|\varphi)p(\varphi)d\varphi} \tag{1.11}$$

assuming that a density on θ exists as well as the conditional density on y.

Of course, we never enjoy an infinite sequence of observations to characterize the prior distribution in practical applications and even if we had these sequences available, the probability measure suggested by de Finetti's theorem could be too complex. Hence, one of the main challenges is to ensure tractable inference procedures with most flexible models. Hence, we have to take a look at how prior knowledge can be incorporated. As mentioned before, this is done by using a prior distribution on the model parameter θ, usually to make predictions about future data.

In the following, we indicate that a quantity may be vector-valued by distinguishing between the notation $\mathbf{y}$ in contrast to y. An important step in Bayesian analysis is the examination of the posterior distribution on $\boldsymbol{\theta}$, given by

$$p(\boldsymbol{\theta}|\mathbf{y},\lambda) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\lambda)}{\int_\Theta p(\mathbf{y}|\boldsymbol{\varphi})p(\boldsymbol{\varphi}|\lambda)d\boldsymbol{\varphi}}, \tag{1.12}$$

where we assume that the necessary densities exist and the n observations are i.i.d. To indicate that the prior distributions are usually parameterized by a set of hyperparameters $\lambda \in \Lambda$ which are usually not the focus of attention. The predicitve likelihood is given by

$$p(\mathbf{y}|\mathbf{y}_1,...,\mathbf{y}_n) = \int_\Theta p(\mathbf{y}|\boldsymbol{\varphi})p(\boldsymbol{\varphi}|\mathbf{y}_1,...,\mathbf{y}_n)d\boldsymbol{\varphi}. \tag{1.13}$$

In specifying the prior distributions, there exist a variety of methods. An objective Bayesian would argue not to include any prior knowledge but to parameterize the prior distribution as flat as possible. The data should speak for itself. This approach leads most prominently to the use of Jeffreys non-informative prior, calculated as the square root of the determinant of the Fisher information, see Jeffreys (1946) or to the reference prior, introduced by Bernardo (1979). A subjective Bayesian would opt for a distribution which represents his subjective prior belief regarding $\boldsymbol{\theta}$. Unfortunately, the integrals of equations (1.12) and (1.13) may be intractable for an arbitrary prior choice. An easy solution are conjugate priors, stemming from the idea that the normalization constant is automatically determined if the posterior distribution is of the same family as the prior distribution and if the functional form of the prior is known.

A prominent example which not only demonstrates the practicability of the Bayesian approach but also hints at a severe weakness in frequentist statistics considers the estimation of the probability of a rare event, e.g. the probability of a company's default in a specific branch. Let π denote the fraction of defaults in the branch and let Y be a random variable denoting the number of defaults, following a binomial probability distribution

$$Y|\pi \quad \sim \quad \mathcal{B}(n, \pi) \tag{1.14}$$

with n observations, where $\mathcal{B}(n, \pi)$ denotes the binomial distribution with the corresponding parameterization. The likelihood is then given as

$$f(y|\pi) \quad = \quad \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i} = \pi^y(1-\pi)^{n-y}, \tag{1.15}$$

where x_i is equal to 1 if firm i defaults and $y = \sum_{i=1}^{n} x_i$ is the number of defaults in the sample. The conjugate prior to the binomial distribution is the Beta-distribution $Beta(a, b)$ and the prior is therefore given as

$$f(\pi) \quad \propto \quad \pi^{-a}(1-\pi)^{-b}. \tag{1.16}$$

The Beta-distribution can easily be calibrated to reflect beliefs regarding π. The expectation for π is given as $\mathbb{E}(\pi) = a/(a+b)$ and the most probable value for π is $(a-1)(a-1+b-1)$. Our uncertainty regarding our beliefs can also be represented in terms of π's variance given as $Var(\pi) = ab/(a+b+1)(a+b)^2$. Choosing a and b could be sufficient to describe the beliefs regarding π sufficiently. After this, calculating the functional form of the posterior distribution is easy, especially due to the similar form of the prior distribution and the likelihood:

$$f(\pi|Y=y) \propto f(y|\pi)f(\pi) \quad \propto \quad \pi^{y-a}(1-\pi)^{n-y-b} \tag{1.17}$$

Hence, $f(\pi|y)$ is the kernel of a $Beta(a+y, b+n-y)$ distribution and the obtained posterior distribution therefore of a well known functional form.

A noteworthy case is $a = b = 0.5$. With such a parameterization, the Beta-distribution is nearly flat and, when used as a prior, would correspond to situation where no prior knowledge should be used in the analysis. Results with such a non-informative prior usually coincide with frequentist approaches, although the mathematical and philosophical foundations are different. Another way to interpret a non-informative prior is having weak beliefs regarding the parameter. A widely used non-informative prior is the Jeffreys-prior, as introduced by Jeffreys (1946). Jeffreys-prior is proportional to the square root of the determinant of the Fisher information matrix. It can also be derived as the limiting case of a conjugate prior, as for example Gelman et al. (2014) point out.

Do we lose something when we decide to use the frequentist or likelihood approach instead? Yes, we do and it is costly. A well known empirical estimator for π is given as $\hat{\pi} = y/n$ which coincides with the Maximum Likelihood estimator. But clearly, $\hat{\pi} = 0$ if we do not observe any defaults in our sample. Of course, this is due to sampling uncertainty which could for example be described by a Wald confidence-interval for a specific confidence level, given as

$$CI(\hat{\pi}) \quad = \quad \hat{\pi} \pm c\sqrt{(\hat{\pi}(1-\hat{\pi}))/n}, \tag{1.18}$$

where c is the $1 - \alpha/2$ quantile of the standard normal distribution for a fixed error rate α. This confidence interval is not an interval for $y = 0$ but a single point $CI(\hat{\pi} = 0) = 0$ and thus failing to describe the sampling uncertainty. A Bayesian approach would lead to a non-zero interval by applying the following steps: Calculate the $\alpha/2$-th and $1 - \alpha/2$-quantile of the posterior distribution, in this case of the $Beta(a+y, b+n-y)$ distribution. But if the posterior distribution is not of a known functional form, a sufficiently large sample from the posterior distribution can be generated and the respective quantiles of the sample can be used instead. Both approaches will lead to an credible interval and not to a single point and thus describes sampling uncertainty more accurately.

This discussion shows that the Bayesian view is a highly flexible approach to empirical questions. This flexibility comes in some cases at almost no costs, especially when using a conjugate prior. And even if no prior knowledge has to be incorporated, the use of a non-informative prior mirrors the frequentist approach but with the advantage of an accessible description of sampling uncertainty.

1.2 A glimpse at modern portfolio theory

Portfolio theory is perhaps one of the most worked on topics at the intersection of mathematics, statistics and economics since Harry Markowitz's 1952 seminal paper. Almost every introduction to portfolio theory explains this huge interest with the nature of a portfolio of assets: everybody has such an allocation of wealth (which is in reality of course not restricted to positive values). And everybody is interested how much her portfolio is worth in the future. This value is equal to the expected value of the sum of the expected returns of all the assets, denoted as $\boldsymbol{\mu}$. But as it is (or should be) common knowledge, a higher $\boldsymbol{\mu}$ is usually related to higher risk. One of the main contributions of Markowitz (1952) is the assertion, that the risk of a portfolio comes from the covariance $\boldsymbol{\Sigma}$ of the assets the portfolio consists of. Hence, portfolio theory typically deals with the trade-off between high returns and low risk. Naturally, an investor is interested in a portfolio with an expected return as high as possible with risk as low as possible, measured in the portfolios's standard deviation. This means on the one hand that an investor would never pic a portfolio with higher risk than another portfolio as long as the expected portfolio returns are the same. On the other hand, an investor would pic the portfolio with the highest expected returns if she has to choose between portfolios with the same risk. This short reflection actually gives the set of efficient portfolios: a portfolio with the highest return among portfolios with the same risk. If there is a risk free asset available, the set of efficient portfolios can be depicted as a straight line, the so called capital market line, otherwise this set is a hyperbola. The tangent point between the capital market line and efficient frontier is called the tangency portfolio. These points are represented in Figure 1.1.

To determine these portfolios, a little more economics is needed. Attitudes towards risk usually differ between investors. While there exists a broad range of research in economics and psychology on how to model risk attitudes, the approach in economics assumes the existence of a utility function in which the investor's preferences of a wealth level (or deviations from it) are displayed. Such a utility function describes the investor's risk preferences as a continuum between risk-aversion and risk-seeking, typically gouverned by a risk-parameter. According to these preferences, the assets of the portfolio are weighted properly. These weights have to be determined. Markowitz (1952) stresses the importance of the covariance of the assets for this purpose. The covariance of the assets actually is the key to a diversification-effect; an effect reducing risk by combining assets which are correlated differently. These assets have to be weighted properly in order to achieve a risk reduction by diversification. Key to find a proper asset weighting is the combination of expected return and covariance of the returns.

Determining these weights depends on several factors: the choice of the utility function is an obvious factor, the market structure is another one. This includes wether or not short selling is permitted, which means the possibility to sell an asset which the seller does not own, if we allow for a risk-free asset and also if more than one period is considered. This has an effect on how the portfolio weights can be determined and has a huge influence on the complexity of the optimization routine to calculate the weights. Clearly, the analytical form of the weights, if

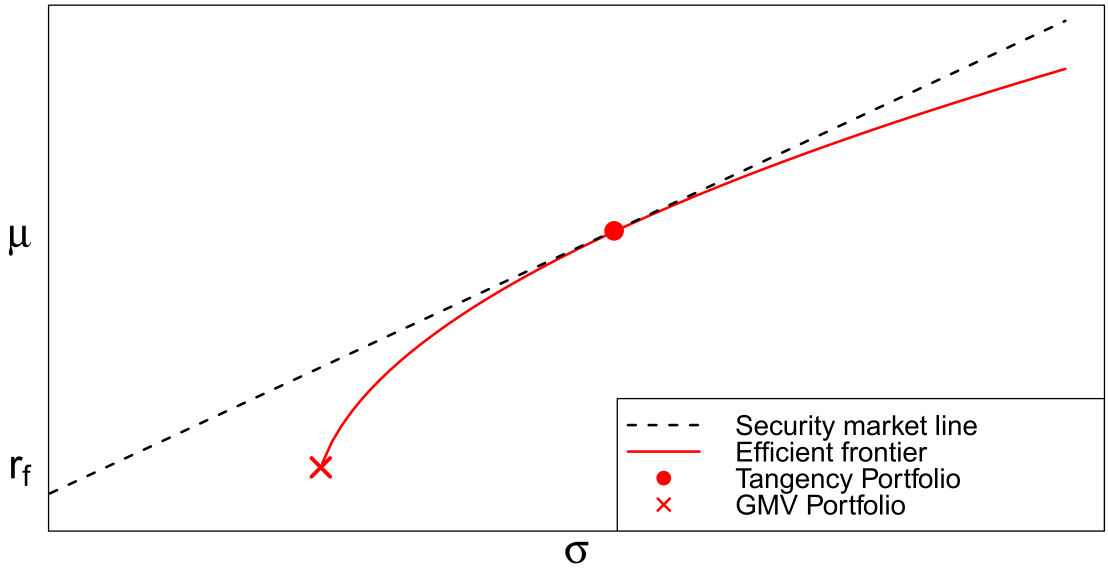

Figure 1.1: The efficient frontier, security market line, Global Minimum Variance portfolio and the Tangent portfolio.
r_f denotes the risk-free interest rate. Chapter 3 deals with the tangent portfolio denoted by the red dot. Chapter 4 concentrates on the whole efficient frontier and the Global Minimum Variance portfolio. In chapter 5, we concentrate mainly on the efficient frontier.

any, can be hugely different and complex. For example, allowing for short sales allows for the possibility of negative weights. Pennacchi (2008), besides many others, provides a vast overview on this topic. A common factor of most of the weights is that most of them are combinations of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For example, the weights of the tangency portfolio are given as

$$\boldsymbol{w}_{TP} = \alpha^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}), \tag{1.19}$$

where α is the coefficient of risk aversion and r_f the return of the risk-free asset. Such products of the precision matrix and the mean vector occur often. The expected return of the global minimum variance portfolio, the portfolio with the least variance among all portfolios, is given as

$$R_{GMV} = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}, \tag{1.20}$$

if we do not allow for a risk-free asset and claim that the sum of all the portfolio weights is equal to one.

When someone speaks of risk in an economic sense, usually uncertainty regarding the future is meant. In practice, the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are replaced by their empirical counterparts and therefore estimated from a dataset. Thus, not only well known economic risk is inherited in the portfolio, but also vast estimation uncertainty. To ignore this estimation risk would not be appropriate when the portfolio's risk has to be described.

1.3 Contributions and outline

The estimation risk of an estimator can statistically be described by its distributional properties. To go back to the Bayesian subsection of this chapter, assume that we have observations from a binomial distribution, $X \sim B(n, \pi)$ and not only want to estimate the success propability π but also to describe the estimation risk of this estimator. Such an estimator is given as

$$\hat{\pi} = \overline{y} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1.21}$$

for n observations $y_1, y_2, ..., y_n$. The distribution of $\hat{\pi}$ is found by an application of the Central Limit Theorem and yields:

$$\hat{\pi} \quad \sim \quad \mathcal{N}(n\pi, \pi(1-\pi)/n). \tag{1.22}$$

Of course, this is well known from every basic statistics course. First of all, this is an approximate result, the sample size has to be sufficiently large. As Brown and DasGupta (2001) showed, this is rather crucial. But a little more intriguing for our purpose here is the right-hand side of (1.22). Obviously, the quantities of the mean's distribution are still not known and therefore not practical. This would look quite differently in Bayesian statistics: the distribution in (1.17) of the parameter π is a $Beta(a + y, b + n - y)$ distribution and contains only empirical or known values. The posterior distribution describes estimation risk directly and enables the practitioner to track this sort of risk in a practical way.

But as the previous section on portfolio theory demonstrates, the parameter expressions in portfolio theory are more complicated. Finding the distribution of combinations of two multivariate random variables $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is a challenge since a convolution of the parameters distribution functions will not result in an analytical expression of the distribution function of this linear combination. We will demonstrate that finding a suitable Bayesian model for the considered portfolios results not only in an accessible posterior distribution for the necessary portfolio parameters, but can also be extended to easily accessible and easily interpretable descriptions of risk, for example a default probability of the portfolio as a whole.

In order to do so, we endow the parameters with their conjugate priors and also with a non-informative prior. Using properties from the theory of multivariate distributions we are able to access the posterior distribution in the form of stochastic representations for the expressions of interest, for example of the portfolio weights or the efficient frontier. The stochastic representations are not only a computationally highly efficient way to sample from the posterior distribution since usual Markov Chain Monte Carlo methods are not needed, but are also the key to determine Bayesian estimates for the parameters.

All of the papers are of the same structure: after an introduction we establish the theory which is illustrated using either a simulation study or an empirical study. After that, the paper concludes. The last section contains the proofs and all supplementary material of the respective paper. The second chapter presents the work of Bauder et al. (2017b). This paper deals with the estimation of a multi-period portfolio with an exponential utility function. Here we established our method for accessing the posterior distribution for the product of a precision matrix and a mean vector. This paper builds upon the solution to the multi-period portfolio given in Bodnar et al. (2015b). This portfolio-model is perhaps the most realistic portfolio considered in this thesis, since investment decisions are usually made for longer than a single period. We derive the stochastic representations for the distribution of the weights and, using this stochastic representations, the Bayesian estimates as well as the variances for the weights. In addition to this, we also state the asymptotic distribution of the weights. To highlight the practical relevance of the portfolio-model as well as the Bayesian approach, we additionally derive the posterior predictive distribution. This allows to calculate the default probability of the portfolio at any point in time. These points are illustrated in an empirical study using data from the FTSE100, covering early summer of 2016. This covers the period of the referendum in the United Kingdom of Great Britain and Northern Island. For this period, we also calculated the default probability using sampled data from the posterior predictive distribution.

The third chapter applies the methods to the tangent portfolio and recapitulates our work in Bauder et al. (2017a). The stochastic representations for the relevant parameters are derived and, in a simulation study, the coverage probabilities of the true posterior distributions and the asymptotic distributions are compared. We find that the coverage probabilities with the diffuse prior almost coincides with the asymptotic distribution. But the conjugate prior shows a better coverage compared to the asymptotic distributions, especially when the returns are strongly correlated. In the fourth chapter, recapitulating Bauder et al. (2018a), we consider the estimation of the whole efficient frontier as well as the parameters of the Global Minimum Variance portfolio. Again, we derived the stochastic representations for the parameters, their

Bayesian point estimates and their asymptotic distributions. We applied the derived expressions to real data, this time to the S&P 500 during the week of the popular vote in the United States in 2016. We mainly focus on credible intervals for the parameters, resulting in a confidence region for the whole efficient frontier and credible sets for the return and variability. Additionally, we were also able to derive the credible intervals for every single portfolio. The fifth and last chapter, as given in Bauder et al. (2018b), deals with a Bayesian approach to mean-variance portfolios in general: here we derive a solution to the portfolio optimization problem which does not depend on unknown quantities. In addition to this, the used posterior predictive distribution also allows to easily construct a prediction interval, similar to the possibilities already demonstrated in the paper on the multi-period portfolio. We compare our method to the standard frequentist approach where it is well known that the slope parameter of the efficient frontier is overoptimistic. We examine the differences which occur especially if the ratio of portfolio dimension to the sample size is moderate to large. Our Bayesian estimator for the efficient frontier is much less overoptimistic. The sixth chapter provides a brief summary and discussion of the previous chapters as well as a brief outline for future research possibilities.

Chapter 2

Bayesian Estimation of the Multi-Period Portfolio for an Exponential Utility

In portfolio theory, the mean-variance paradigm introduced by Markowitz (1952) is still a popular reference for understanding the relationship between systematic risk, return and investment behaviour. A portfolio is determined here by using the asset expected returns and their covariances. As a starting point, Markowitz (1952) was vastly extended in the following 70 years. While Markowitz (1952) focused only on a single investment period, the multi-period solution was introduced in Markowitz (1959). Merton (1969) showed that the mean-variance multi-period setting in the continuous time case is equivalent to expected utility maximization for an exponential utility function. The multi-period optimal portfolio choice problems for different utility functions were considered by Mossin (1968), Samuelson (1969), Elton (1974), Brandt and Santa-Clara (2006), Basak and Chabakauri (2010).

While these studies focus on the continuous time case, Li and Ng (2000), Çanakoğlu and Özekici (2009), Bodnar et al. (2015a,b) presented the results in the discrete time case for the quadratic utility function and the exponential utility function. In particular, Bodnar et al. (2015b) derived an analytical expression for the multi-period optimal portfolio weights under the assumption of non-tradable predictable variables and a VAR(1)-structure which are described as linear combinations of the precision matrix (inverse covariance matrix) and the expected return vector. While this setting allows for flexibility in building trading strategies under quite unrestrictive assumptions, there are still shortcomings: (i) since the parameters of the asset return distribution, namely the mean vector and the covariance matrix, are unknown quantities, the optimal portfolio weights cannot be constructed in practice and they are obtained by replacing the unknown parameter of the asset return distribution by the corresponding estimates; (ii) although the distributional properties of the estimated optimal portfolio weights and corresponding inference procedures were derived in a number of literature studies for the single-period investment strategies (see, e.g., Gibbons et al. (1989), Shanken (1992), Shanken and Zhou (2007), Okhrin and Schmid (2006), Bodnar and Schmid (2008a, 2011), Bodnar and Schmid (2009)), the problem with the overlapping estimation windows appears to be very crucial under the multi-period setting; (iii) due to the multivariate structure, the determination of the joint distribution of the estimated multi-period optimal portfolio weights is a challenging task.

To tackle all these three challenges, we opt for a Bayesian approach. The Bayesian approach is a well established method for building trading strategies in a single-period optimal portfolio choice problem, starting with Winkler (1973) and Winkler and Barry (1975) and continued until this day. For an overview, see, e.g., Brandt (2010) where also Bayesian portfolio methods are discussed, or Avramov and Zhou (2010). As Avramov and Zhou (2010) pointed out, the Bayesian setting is a realistic description of human decision making processes and information utilization. Both past events and experiences influence the beliefs of market participants at least up to a certain degree how an investment will develop. The investor beliefs are modeled via a prior distributions which represents the relevant information regarding the behaviour of the asset returns. While there is a plenty of possibilities to specify the prior, we focus on the non-informative diffuse prior and the informative conjugate prior (see, e.g., Zellner (1971), and Gelman et al. (2014)) not only for computational reasons but mainly because of their popularity in the financial literature (c.f., Barry (1974), Brown (1976), Klein and Bawa (1976), Frost and Savarino (1986), Aguilar and West (2000), Rachev et al. (2008), Avramov and Zhou (2010), Sekerke (2015), Bodnar et al. (2017b)). Furthermore, their application allows to derive the corresponding posterior distributions in the closed-form what enables us to access important risk measures and to construct credible sets.

The obtained posterior distributions of the optimal portfolio weights under both employed priors are presented in terms of their stochastic representations. A stochastic representation is a well established tool in computational statistics (c.f., Givens and Hoeting (2012)) and in the theory of elliptically contoured distributions (see, e.g. Gupta et al. (2013)) which was already used in Bayesian statistics by Bodnar et al. (2017b). It turns out that the derived stochastic representations are very powerful, allowing us to access not only the posterior distribution of the multi-period optimal portfolio weights, but also to determine the predictive distribution for the wealth at each point of the holding period. Therefore, we are able to access the quantiles for the posterior predictive wealth distribution and can calculate the risk associated with the portfolio at every point over the lifetime of a portfolio, besides analytical Bayesian estimates for the weights together with their uncertainties. Besides these pleasing properties, the developed stochastic representations are highly efficient from a computational point of view since Markov-Chain Monte-Carlo methods are no longer needed. In addition to the derivation of these results, we illustrate this method and its properties on real data. We test the model in an exhaustive study using data from the FTSE 100, where the portfolios cover the time of Great Britains referendum to leave the European Union on 23.6.2016, more commonly regarded as "Brexit", where a slim majority of British voters decided to leave the European Union. Although this result was regarded as the less likely option in advance, it was regarded as the option with the least favourable effects on the British economy and should therefore have an effect on a portfolio covering this period.

The remaining chapter is structured in the following way. In section 2.1.1, we briefly review the solution of the multi-period optial portfolio choice problem with exponential utility derived in Bodnar et al. (2015b). The stochastic representations for the optimal portfolio weights under both priors are presented in Theorems 3 and 4 (section 2.1.2), which are use to derive the corresponding Bayes estimates for the weights (Theorem 5) together with their covariance matrix (Theorem 6) as well as to prove the posterior asymptotic normality (Theorem 7). In section 2.1.3, we obtain the posterior predictive distribution for the wealth during the holding period which is provided in terms of stochastic representation in Theorem 8 under both employed priors. In section 2.2, the suggested Bayesian approach is applied to the Brexit-data by calculating the asymptotic distributions for the optimal portfolio weights, determining the credible sets for the portfolio wealth and specifying the default probabilities at each time point. Section 2.3 summarizes the main results of the chapter, while all technical proofs are moved to the appendix to this chapter (section 2.4).

2.1 Bayesian analysis of multi-period optimal portfolios

2.1.1 Analytical solution of the multi-period optimization problem

Let $\mathbf{X}_t = (X_{t,1}, X_{t,2}, ..., X_{t,k})^\top$ be a random vector of returns on k assets taken at time point t. Throughout the paper we assume that the asset returns $\mathbf{X}_1, \mathbf{X}_2, ...$ are infinitely exchangeable and multivariate centered spherically symmetric. This assumption, in particular, implies (see, e.g., Bernardo and Smith (2000, Proposition 4.6)) that the asset returns are independently and identically distributed given the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ with the conditional distribution given by $\mathbf{X}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (k-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$). It is noted that the imposed assumption imply that neither the unconditional distribution of the asset returns is normal nor that they are independently distributed. Moroever, the unconditional distribution of the asset returns appears to be heavy-tailed which is usually observed for financial data.

The quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the parameters of the asset returns distribution where $\boldsymbol{\Sigma}$ is assumed to be a $k \times k$ dimensional positive definite matrix. We consider a multi-period portfolio choice problem with the allocation of initial wealth at time point $t = 0$ and with the subsequent update of the portfolio structure at time points $t \in \{1, 2, ..., T\}$. Let $\mathbf{v}_t = (v_{t,1}, ..., v_{t,k})^\top$ stand for the vector of portfolio weights determined at time t and let $r_{f,t}$ be the return on the risk-free asset in period t. We assume that short-selling is allowed, i.e. the weights could also be negative. The vector $\mathbf{v}_t$ specifies the structure of the portfolio related to the risky assets, whereas the part of the wealth equal to $1 - \mathbf{1}^\top \mathbf{v}_t$ is invested into the risk-free asset where $\mathbf{1}$ denotes the k-dimensional vector of ones. Then the investor's wealth in period t is expressed as

$$W_t = W_{t-1}(1 + (1 - \mathbf{1}^\top \mathbf{v}_{t-1}) r_{f,t} + \mathbf{v}_{t-1}^\top \mathbf{X}_t) = W_{t-1}(1 + r_{f,t} + \mathbf{v}_{t-1}^\top (\mathbf{X}_t - r_{f,t}\mathbf{1})).$$

An investor seeks to maximize the utility of the final wealth, i.e. $U(W_T)$, where $U(x) = -\exp(-\gamma x)$ is the exponential utility function and the coefficient of absolute risk aversion, $\gamma > 0$, determines the investor's attitude towards risk. The optimization problem is given by

$$V(0, W_0) = \max_{\{\mathbf{v}_s\}_{s=0}^{T-1}} \mathbb{E}_0[U(W_T)] \tag{2.1}$$

where the maximum is taken with respect to all weights $\mathbf{v}_0$,..., $\mathbf{v}_{T-1}$ which specify the portfolio structure during the initial period of investment as well as during all consequent reallocations. The solution of (2.1) is derived in the recursive way starting from the last period by applying Bellman equations at 0, 1, ... $T-1$. The optimization problem at time point $T-t$ is then given by

$$\begin{aligned} V(T-t, W_{T-t}) &= \max_{\{\mathbf{v}_s\}_{s=T-t}^{T-1}} \mathbb{E}_{T-t}\left[\max_{\{\mathbf{v}_s\}_{s=T-t+1}^{T-1}} \mathbb{E}_{T-t+1}[U(W_T)]\right] \\ &= \max_{\mathbf{v}_{T-t}} \mathbb{E}_{T-t}\left[V(T-t+1, W_{T-t}\left(r_{f,T-t} + \boldsymbol{w}_{T-t+1}^\top(\mathbf{X}_{T-t+1} - r_{f,T-t+1}\mathbf{1})\right))\right] \end{aligned}$$

subject to the terminal condition $U(W_T) = -\exp(-\gamma W_T)$ with $\boldsymbol{w}_{T-t+1}$ as the optimal portfolio weights in period $T-t+1$. For details on this method, see e.g. Pennacchi (2008), while Bodnar et al. (2015b) determine an analytical solution of (2.1) under the exponential utility. The latter results are summarized in Proposition 1.

Proposition 1. *Let* $\mathbf{X}_t$, $t = 0, ..., T$ *be a sequence of conditionally independently and identically distributed vectors of* k *risky assets with* $\mathbf{X}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Let* $\boldsymbol{\Sigma}$ *be positive definite. Then* *the optimal multi-period portfolio weights are given by*

$$\boldsymbol{w}_t = C_t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_{f,t+1}\mathbf{1}), \quad \textit{with} \quad C_t = (\gamma W_t \prod_{i=t+2}^{T} R_{f,i})^{-1} \tag{2.2}$$

for $t = 0, ..., T-1$ *where* $R_{f,i} = 1 + r_{f,i}$ *and* $\prod_{i=T+1}^{T} R_{f,i} \equiv 1$.

Although Proposition 1 provides a simple solution of the multi-period portfolio choice problem, the formula (2.2) cannot directly be applied in practice since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown parameters of the asset return distribution. As a result, these two quantities have to be estimated before the portfolio (2.2) is constructed. However, the usage the estimated mean vector and the estimated covariance matrix instead of the population ones does not ensure that the estimated portfolio weights coincide with true ones. Then two main questions raise: (i) how strongly deviates the estimated portfolio from the population one? and (ii) is it reasonable to invest into the estimated portfolio? Both questions have to be treated by using statistical methods and are very closely connected to the distributional properties of the estimates constructed for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The traditional approach of estimating the portfolio weights relies on the methods from the conventional statistics where the sample mean vector and the sample covariance matrix are used. Let $\mathbf{x}_{t-n+1}, ..., \mathbf{x}_t$ be the observation vectors of asset returns which are considered as realizations of the corresponding random vectors $\mathbf{X}_i$, $i = t-n+1, ..., t$. Then the mean vector and the covariance matrix at time point t are estimated by

$$\overline{\mathbf{x}}_t = \frac{1}{n} \sum_{i=t-n+1}^{t} \mathbf{x}_i \quad \text{and} \quad \mathbf{S}_t = \frac{1}{n-1} \sum_{i=t-n+1}^{t} (\mathbf{x}_i - \overline{\mathbf{x}}_t)(\mathbf{x}_i - \overline{\mathbf{x}}_t)^\top . \tag{2.3}$$

The sample estimate of the multi-period optimal portfolio is obtained by replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2.2) by the corresponding estimates from (2.3). This leads to

$$\hat{\boldsymbol{w}}_t = C_t \mathbf{S}_t^{-1}(\overline{\mathbf{x}}_t - r_{f,t+1}\mathbf{1}) \quad \text{with} \quad C_t = (\gamma W_t \prod_{i=t+2}^{T} R_{f,i})^{-1} \quad \text{for} \quad t = 0, ..., T-1. \tag{2.4}$$

Using the findings in Bodnar and Okhrin (2011), we obtain the density function, the moments and the stochastic representation of the sample multi-period optimal portfolio weights from the viewpoint of frequentist statistics. These results provide answers on the above two questions and allow us to characterize the distributional properties of each vector of weights $\hat{\boldsymbol{w}}_t$ separately. On the other hand, they do not take into account the multi-period nature of the considered investment procedure. More precisely, it is not possible to provide the characterization of the whole multi-period optimal portfolio, since the overlapping samples are used and the dependence structure between the estimated portfolio weights becomes severe.

For that reason, we deal with the problem of estimating the multi-period optimal portfolio from the viewpoint of Bayesian statistics and consider the portfolio constructed by using (2.4) as a benchmark portfolio without investigating its distributional properties in detail. In contrast to the methods of the frequentist statistics, the application of the Bayesian approach allows the sequential update of the available information which is a very important property needed for estimating the multi-period portfolio weights.

2.1.2 Bayesian estimation of portfolio weights

Let $\mathbf{x}_{t,n} = (\mathbf{x}_{t-n+1}, ..., \mathbf{x}_t)$ denote the observation matrix at time point t which consists of n asset return vectors from $t-n+1$ to t. According to Bayes theorem, the beliefs regarding $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are updated in the presence of occurring data, yielding the posterior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}_{t,n})$ to be proportional to the product of the likelihood function $L(\mathbf{x}_{t,n}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the prior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The posterior is, then, used to derive Bayesian estimates for the multi-period optimal portfolio weights as well as their characteristics, like the covariance matrix and a credible region which is an analogue to a confidence region in the conventional statistics. The Bayes theorem states that

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}_{t,n}) \quad \propto \quad L(\mathbf{x}_{t,n}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The choice of the prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is an important step in the Bayesian decision process. Although the prior should reflect the investor's belief regarding the parameters of the asset return distribution, it also strongly affects the model's computational properties since it influences the accessibility of the posterior distribution. Several priors for the mean vector and covariance matrix of the asset returns have been suggested in literature (see, e.g., Barry (1974), Brown (1976), Klein and Bawa (1976), Frost and Savarino (1986), Rachev et al. (2008), Avramov and Zhou (2010), Sekerke (2015)) with the recent paper of Bodnar et al. (2017b) summarizing these results. In the following, we choose Jeffreys' non-informative prior and a conjugate informative prior for both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. These two priors are widely used in the context of Bayesian inference of optimal portfolios.

The Jeffreys non-informative prior, also known as the diffuse prior, is given by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \propto \quad |\boldsymbol{\Sigma}|^{-(k+1)/2} \tag{2.5}$$

while the cojugate prior is expressed as

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathcal{N}_k\left(\mathbf{m}_0, \frac{1}{r_0}\boldsymbol{\Sigma}\right), \tag{2.6}$$
$$\boldsymbol{\Sigma} \sim \mathcal{IW}_k(d_0, \mathbf{S}_0), \tag{2.7}$$

where $\mathbf{m}_0$, r_0, d_0, $\mathbf{S}_0$ are additional model parameters known as hyperparameters. The symbol $\mathcal{IW}_k(d_0, \mathbf{S}_0)$ denotes the inverse Wishart distribution with d_0 degrees of freedom and parameter matrix $\mathbf{S}_0$. The prior mean $\boldsymbol{\mu}_0$ reflects our prior expectations about the expected asset returns, while $\mathbf{S}_0$ presents in the model the prior beliefs about the covariance matrix. The other two hyperparameters r_0 and d_0 are known as precision parameters for $\boldsymbol{\mu}_0$ and $\mathbf{S}_0$, respectively. Note that the prior (2.6)-(2.7) corresponds to the well-known conjugate normal-inverse-Wishart model as discussed by, e.g., Gelman et al. (2014). In this case the posterior is accessible in an analytical form and moreover, has the same distribution as the prior with updated hyperparameters.

In Proposition 2, we present the marginal posterior of $\boldsymbol{\mu}$ as well as the conditional posterior of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$. These results will be later used in the derivation of Bayesian estimates for the optimal portfolio weights. In the following the symbol $t_k(d, \mathbf{a}, \mathbf{A})$ stands for the multivariate k-dimensional t-distribution with d degrees of freedom, location vector $\mathbf{a}$ and dispersion matrix $\mathbf{A}$. In the case of $k = 1$, $\mathbf{a} = 0$, and $\mathbf{A} = 1$, we use the notation t_d to denote the standard univariate t-distribution with d degrees of freedom.

Proposition 2. *Let $\mathbf{X}_{t-n+1}, ..., \mathbf{X}_t$ be conditionally independently distributed with $\mathbf{X}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = t-n+1, ..., t$ with $n > k$. Then:*

(a) Under the diffuse prior (2.5), the marginal posterior distribution of $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu}|\mathbf{x}_{t,n} \sim t_k\left(n-k, \overline{\mathbf{x}}_{t,d}, \frac{1}{n(n-k)}\mathbf{S}_{t,d}\right) \quad \text{with} \quad \overline{\mathbf{x}}_{t,d} = \overline{\mathbf{x}}_t \quad \text{and} \quad \mathbf{S}_{t,d} = (n-1)\mathbf{S}_t.$$

The conditional posterior distribution of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ is expressed as

$$\boldsymbol{\Sigma}|\boldsymbol{\mu}, \mathbf{x}_{t,n} \sim \mathcal{IW}_k(n+k+1, \mathbf{S}^*_{t,d}(\boldsymbol{\mu})) \quad \text{with} \quad \mathbf{S}^*_{t,d}(\boldsymbol{\mu}) = \mathbf{S}_{t,d} + n(\boldsymbol{\mu} - \overline{\mathbf{x}}_{t,d})(\boldsymbol{\mu} - \overline{\mathbf{x}}_{t,d})^\top.$$

(b) Under the conjugate prior (2.6) and (2.7), the marginal posterior distribution of $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu}|\mathbf{x}_{t,n} \sim t_k\left(n+d_0-2k, \overline{\mathbf{x}}_{t,c}, \frac{1}{(n+r_0)(n+d_0-2k)}\mathbf{S}_{t,c}\right) \quad \text{with}$$

$$\overline{\mathbf{x}}_{t,c} = \frac{n\overline{\mathbf{x}}_t + r_0\mathbf{m}_0}{n + r_0} \quad \textit{and} \quad \mathbf{S}_{t,c} = \mathbf{S}_{t,d} + \mathbf{S}_0 + nr_0\frac{(\mathbf{m}_0 - \overline{\mathbf{x}}_{t,c})(\mathbf{m}_0 - \overline{\mathbf{x}}_{t,c})^\top}{n + r_0}.$$

The conditional posterior distribution of $\boldsymbol{\Sigma}$ *given* $\boldsymbol{\mu}$ *is expressed as*

$$\begin{aligned} \boldsymbol{\Sigma}|\boldsymbol{\mu}, \mathbf{x}_{t,n} &\sim IW_k(n + d_0 + 1, \mathbf{S}^*_{t,c}(\boldsymbol{\mu})) \quad \textit{with} \\ \mathbf{S}^*_{t,c}(\boldsymbol{\mu}) &= \mathbf{S}_{t,c} + (n + r_0)(\boldsymbol{\mu} - \overline{\mathbf{x}}_{t,c})(\boldsymbol{\mu} - \overline{\mathbf{x}}_{t,c})^\top. \end{aligned}$$

The proof of Proposition 2 follows from chapter 3 in Gelman et al. (2014) who presented the expressions of the marginal posterior distributions of $\boldsymbol{\mu}$ under both the diffuse and the conjugate priors. Then, the results for the conditional posteriors of $\boldsymbol{\Sigma}$ are obtained from the joint posterior distributions using the formulae for the marginal posteriors for $\boldsymbol{\mu}$. It is remarkable that although the results for the marginal posteriors for both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are widely used in Bayesian inferences and the conditional posteriors for $\boldsymbol{\mu}$ given $\boldsymbol{\Sigma}$ have been considered previously in literature (see, e.g., Sun and Berger (2007)), the results for the conditional posteriors of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ have not been discussed nor used. Next, we show that the last finding allows to derive posterior distributions for functions which includes both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

In order to assess the risk associated with estimating the optimal portfolio weights, we need to derive results about the posterior distribution of the weights presented in Proposition 1 which are given as a product of the inverse covariance matrix and the mean vector. Next, we establish very useful stochastic representations for these weights, endowing the parameters with their diffuse and conjugate priors. The results are summarized in Theorem 3, where the stochastic representations are derived for an arbitrary linear combination of optimal portfolio weights. These findings are later used for calculating the Bayesian estimates of the portfolio weights (Theorem 5) and their covariance matrix (Theorem 6). It is noted that the application of the stochastic representation to describe the distribution of random quantities has been used both in the conventional statistics (see, e.g., Givens and Hoeting (2012), Gupta et al. (2013)) and the Bayesian statistics (c.f., Bodnar et al. (2017b)). Later on, the symbol "$\overset{d}{=}$" denotes the equality in distribution. The proof of Theorem 3 is presented in section 2.4.

Theorem 3. *Let* $\mathbf{L}$ *be a* $p \times k$*-dimensional matrix of constants. Then under the assumption of Proposition 2 we get:*

(a) Under the diffuse prior (2.5), the stochastic representation of $\mathbf{L}\boldsymbol{w}_t$ *is given by*

$$\begin{aligned} \mathbf{L}\boldsymbol{w}_t &\overset{d}{=} C_t\eta\mathbf{L}\mathbf{S}^*_{t,d}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) + C_t\sqrt{\eta}\Big((\boldsymbol{\mu} - r_{f,t+1})^\top\mathbf{S}^*_{t,d}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) \cdot \mathbf{L}\mathbf{S}^*_{t,d}(\boldsymbol{\mu})^{-1}\mathbf{L}^\top \\ &- \mathbf{L}\mathbf{S}^*_{t,d}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})(\boldsymbol{\mu} - r_{f,t+1})^\top\mathbf{S}^*_{t,d}(\boldsymbol{\mu})^{-1}\mathbf{L}^\top\Big)^{1/2}\mathbf{z}_0, \end{aligned}$$

where $\eta \sim \chi^2_n$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, *and* $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n-k, \overline{\mathbf{x}}_{t,d}, \mathbf{S}_{t,d}/(n(n-k))\right)$; *moreover,* $\eta, \mathbf{z}_0$ *and* $\boldsymbol{\mu}$ *are mutually independent.*

(b) Under the conjugate prior (2.6) *and* (2.7), *the stochastic representation of* $\mathbf{L}\boldsymbol{w}_t$ *is given by*

$$\begin{aligned}\mathbf{L}\boldsymbol{w}_t \overset{d}{=}\; & C_t\eta\mathbf{L}\mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu}-r_{f,t+1}) + C_t\sqrt{\eta}\Big((\boldsymbol{\mu}-r_{f,t+1})^\top\mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu}-r_{f,t+1})\cdot\mathbf{L}\mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}\mathbf{L}^\top \\ & - \mathbf{L}\mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu}-r_{f,t+1})(\boldsymbol{\mu}-r_{f,t+1})^\top\mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}\mathbf{L}^\top\Big)^{1/2}\mathbf{z}_0,\end{aligned}$$

where $\eta \sim \chi^2_{n+d_0-k}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, *and* $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n+d_0-2k, \overline{\mathbf{x}}_{t,c}, \mathbf{S}_{t,c}/((n+r_0)(n+d_0-2k))\right)$; *moreover,* $\eta, \mathbf{z}_0$ *and* $\boldsymbol{\mu}$ *are mutually independent.*

The results of Theorem 3 show that in both cases, i.e., when the mean vector and the covariance matrix are endowed by the diffuse prior and the conjugate prior, the obtained stochastic representations are very similar and the posterior distributions of the multi-period optimal portfolio weights from Proposition 1 can be described by three random variables which have standard univariate/multivariate distributions.

Another important application of Theorem 3 is that the results of this theorem also provide a hint how these distributions can be accessed in practice via simulations, namely by simulating samples from the χ^2-distribution, the normal distribution, and the t-distribution. Although the derived stochastic representations have some nice computational properties in terms of speed, they are not computationally efficient. In the following theorem we derive further stochastic representations under both priors by applying the Sherman-Morrison-Woodbury formula on the inverse of the posterior scale matrices $\mathbf{S}^*_{t,d}(\boldsymbol{\mu})$ and $\mathbf{S}^*_{t,c}(\boldsymbol{\mu})$. The proof of the theorem is provided in the appendix. Let $\mathcal{F}(d_1, d_2)$ denote the F-distribution with d_1 and d_2 degrees of freedom.

Theorem 4. *Under the assumption of Theorem 3 we get:*

(a) Under the diffuse prior (2.5), the stochastic representation of $\mathbf{L}\boldsymbol{w}_t$ *is given by*

$$\mathbf{L}\boldsymbol{w}_t \overset{d}{=} C_t\eta\mathbf{L}\boldsymbol{\zeta}_d + C_t\sqrt{\eta}\left(\epsilon_d\mathbf{L}\boldsymbol{\Upsilon}_d\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_d\boldsymbol{\zeta}_d^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{2.8}$$

with

$$
\begin{aligned}
\epsilon_d &= \epsilon_d(Q,\mathbf{u}) = (\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})\\
&+ \frac{2}{\sqrt{n}}\frac{\sqrt{kQ/(n-k)}}{1+kQ/(n-k)}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}\\
&+ \frac{1}{n}\frac{kQ/(n-k)}{1+kQ/(n-k)} - \frac{kQ/(n-k)}{1+kQ/(n-k)}\left((\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}\right)^2,\\
\boldsymbol{\zeta}_d &= \boldsymbol{\zeta}_d(Q,\mathbf{u}) = \mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) + \frac{1}{\sqrt{n}}\frac{\sqrt{kQ/(n-k)}}{1+kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\mathbf{u}\\
&- \frac{kQ/(n-k)}{1+kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,d}^{-1/2}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}),\\
\boldsymbol{\Upsilon}_d &= \boldsymbol{\Upsilon}_d(Q,\mathbf{u}) = \mathbf{S}_{t,d}^{-1} - \frac{kQ/(n-k)}{1+kQ/(n-k)}\mathbf{S}_{t,d}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,d}^{-1/2},
\end{aligned}
$$

where $\eta \sim \chi^2_n$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, $Q \sim \mathcal{F}(k, n-k)$, *and* $\mathbf{u}$ *uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η, $\mathbf{z}_0$, Q, *and* $\mathbf{u}$ *are mutually independent.*

(b) *Under the conjugate prior* (2.6) *and* (2.7)*, the stochastic representation of* $\mathbf{L}\boldsymbol{w}_t$ *is given by*

$$
\mathbf{L}\boldsymbol{w}_t \overset{d}{=} C_t\eta\mathbf{L}\boldsymbol{\zeta}_c + C_t\sqrt{\eta}\left(\epsilon_c\mathbf{L}\boldsymbol{\Upsilon}_c\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_c\boldsymbol{\zeta}_c^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{2.9}
$$

with

$$
\begin{aligned}
\epsilon_c &= \epsilon_d(Q,\mathbf{u}) = (\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})\\
&+ \frac{2}{\sqrt{n+r_0}}\frac{\sqrt{kQ/(n+d_0-2k)}}{1+kQ/(n+d_0-2k)}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}\\
&+ \frac{1}{n+r_0}\frac{kQ/(n+d_0-2k)}{1+kQ/(n+d_0-2k)}\\
&- \frac{kQ/(n+d_0-2k)}{1+kQ/(n+d_0-2k)}\left((\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1/2}\mathbf{u}\right)^2,\\
\boldsymbol{\zeta}_c &= \boldsymbol{\zeta}_d(Q,\mathbf{u}) = \mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}) + \frac{1}{\sqrt{n+r_0}}\frac{\sqrt{kQ/(n+d_0-2k)}}{1+kQ/(n+d_0-2k)}\mathbf{S}_{t,c}^{-1/2}\mathbf{u}\\
&- \frac{kQ/(n+d_0-2k)}{1+kQ/(n+d_0-2k)}\mathbf{S}_{t,c}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,c}^{-1/2}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}),\\
\boldsymbol{\Upsilon}_c &= \boldsymbol{\Upsilon}_d(Q,\mathbf{u}) = \mathbf{S}_{t,c}^{-1} - \frac{kQ/(n+d_0-2k)}{1+kQ/(n+d_0-2k)}\mathbf{S}_{t,c}^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_{t,c}^{-1/2},
\end{aligned}
$$

where $\eta \sim \chi^2_{n+d_0-k}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, $Q \sim \mathcal{F}(k, n+d_0-2k)$, *and* $\mathbf{u}$ *uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η, $\mathbf{z}_0$, Q, *and* $\mathbf{u}$ *are mutually independent.*

Theorem 4 provides alternative stochastic representations of the optimal portfolio weights obtained under the diffuse prior and under the conjugate prior. Although more difficult mathematical expressions are present in Theorem 4, they are more computationally efficient than the ones provided in Theorem 3. Namely, there is no need to calculate the inverse of the matrices $\mathbf{S}^*_{t,d}(\boldsymbol{\mu})$ and $\mathbf{S}^*_{t,c}(\boldsymbol{\mu})$ in each simulation run and instead, we only calculate the inverse of the matrices $\mathbf{S}_{t,d}$ and $\mathbf{S}_{t,c}$ once for the whole simulation study. This property surely speeds up the simulation study considerably. Finally, we note that the realizations of the random vector $\mathbf{u}$, which is uniformly distributed on the unit sphere in $\mathbb{R}^k$, are obtained by drawing $\mathbf{z}$ from the k-dimensional standard normal distribution and calculating $\mathbf{u} = \mathbf{z}/\sqrt{\mathbf{z}^\top \mathbf{z}}$.

The results of Theorem 4 are used to derive Bayesian estimates for the weights of the multi-period optimal portfolio at the initial period of investment as well as at each time of reallocations. They are presented in Theorem 5.

Theorem 5. *Under the assumption of Theorem 3, we get*

(a) Under the diffuse prior (2.5), the Bayes estimate for the optimal portfolio weights at time point t is given by

$$\hat{\boldsymbol{w}}_{t,d} = \mathbb{E}(\boldsymbol{w}_t|\mathbf{x}_{t,n}) \quad = \quad C_t(n-1)\mathbf{S}^{-1}_{t,d}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})\,.$$

(b) Under the conjugate prior (2.6) *and* (2.7)*, the Bayes estimate for the optimal portfolio weights at time point t is given by*

$$\hat{\boldsymbol{w}}_{t,c} = \mathbb{E}(\boldsymbol{w}_t|\mathbf{x}_{t,n}) \quad = \quad C_t(n+d_0-k-1)\mathbf{S}^{-1}_{t,c}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})\,.$$

The proof of the theorem is given in section 2.4. It is interesting to note that the estimate for the optimal portfolio weights obtained under the diffuse prior coincides with the expression derived in Section 2.1.2 for their frequentist estimate since $\mathbf{S}_{t,d}/(n-1) = \mathbf{S}_t$.

Finally, we present the expressions for the covariance matrices of the optimal portfolio weights in Theorem 6 with the proof moved to the appendix. These formulas characterize the dependencies between the portfolio weight and also allow to access their Bayesian risk.

Theorem 6. *Under the assumption of Theorem 3, we get:*

(a) Under the diffuse prior (2.5), the covariance matrix of $\boldsymbol{w}_t$ is given by

$$\begin{aligned}\mathbf{V}_{t,d} = \mathbb{V}ar(\boldsymbol{w}_t|\mathbf{x}_{t,n}) &= C_t^2\Bigg[(n-1)\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1}\\ &+ \left(\frac{n^2+k-2}{n(n+2)} + \frac{k-1}{k}b_d\right)\mathbf{S}_{t,d}^{-1}\Bigg],\end{aligned}$$

where $b_d = n(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$.

(b) Under the conjugate prior (2.6) and (2.7), the covariance matrix of $\boldsymbol{w}_t$ is given by

$$\begin{aligned}\mathbf{V}_{t,c} = \mathbb{V}ar(\boldsymbol{w}_t|\mathbf{x}_{t,n}) &= C_t^2\Bigg[(n+d_0-k-1)\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1}\\ &+ \left(\frac{(n+d_0-k)^2+k-2}{(n+r_0)(n+d_0-k+2)} + \frac{(n+d_0-k)(k-1)}{(n+r_0)k}b_c\right)\mathbf{S}_{t,c}^{-1}\Bigg],\end{aligned}$$

where $b_c = (n+r_0)(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$.

The results of Theorems 5 and 6 provide the first two moments of optimal portfolio weights and, consequently, they characterize their mean values, variances, and correlations. Although different formulas are obtained under the diffuse prior and under the conjugate prior, when the sample size increases the difference between the corresponding expressions becomes negligible.

More general results are provided in Theorem 7 where it is shown that $\boldsymbol{w}_t$ converge to the same asymptotic normal distribution under the diffuse prior and under the conjugate prior.

Theorem 7. *Under the assumption of Theorem 3, it holds that*

$$\begin{aligned}\sqrt{n}(\boldsymbol{w}_t - \hat{\boldsymbol{w}}_t)|\mathbf{x}_{t,n} &\xrightarrow{d} \mathcal{N}\Bigg(0, C_t^2\Bigg[\breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top\breve{\mathbf{S}}_t^{-1}\\ &+ \left(1 + \frac{k-1}{k}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top\breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})\right)\breve{\mathbf{S}}_t^{-1}\Bigg]\Bigg)\end{aligned}$$

as $n \longrightarrow \infty$ under both the diffuse prior and the conjugate prior where

$$\breve{\mathbf{x}}_t \equiv \lim_{n\longrightarrow\infty}\overline{\mathbf{x}}_{t,d} = \lim_{n\longrightarrow\infty}\overline{\mathbf{x}}_{t,c} \quad \textit{and} \quad \breve{\mathbf{S}}_t \equiv \lim_{n\longrightarrow\infty}\frac{\mathbf{S}_{t,d}}{n-1} = \lim_{n\longrightarrow\infty}\frac{\mathbf{S}_{t,c}}{n+r_0}$$

and

$$\hat{\boldsymbol{w}}_t \equiv \lim_{n\longrightarrow\infty}\hat{\boldsymbol{w}}_{t,d} = \lim_{n\longrightarrow\infty}\hat{\boldsymbol{w}}_{t,c} = C_t\breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1}).$$

The proof of Theorem 7 is given in the appendix. Its results are in line with the Bernstein-von Mises theorem (c.f., Bernardo and Smith (2000)) which shows under some regularity conditions that the posterior distribution converges to the normal one independently of the prior used when the sample size tends to infinity. In practice, the asymptotic covariance matrix of $\boldsymbol{w}_t$ is approximated by using $\overline{\mathbf{x}}_t$ and $\mathbf{S}_t$ instead of $\breve{\mathbf{x}}_t$ and $\breve{\mathbf{S}}_t$.

2.1.3 Posterior predictive distribution

In this section we derive the posterior predictive distribution of the wealth at time point $t+1$, $\widehat{W}_{t+1}$, given the observable data $\mathbf{x}_{t,n}$ under the diffuse prior (2.5) and the conjugate prior(2.6) and (2.7) for the given vector of portfolio weights $\mathbf{v}_t$ and the current wealth W_t. Namely, the aim is to derive the posterior predictive distribution of

$$W_{t+1} = W_t(1 + r_{f,t} + \mathbf{v}_t^\top(\mathbf{X}_{t+1} - r_{f,t+1})) \tag{2.10}$$

given information provided by the observation matrix $\mathbf{x}_{t,n}$, i.e.

$$f_{\hat{W}_{t+1}}(w|\mathbf{x}_{t,n}) \quad = \quad \int_{\boldsymbol{\mu},\boldsymbol{\Sigma}} f_{\hat{W}_{t+1}}(w|\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{x}_{t,n})\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{x}_{t,n})\mathrm{d}\boldsymbol{\mu}\mathrm{d}\boldsymbol{\Sigma},$$

where $\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{x}_{t,n})$ is the posterior distribution obtained under the diffuse prior or the conjugate prior. The symbol $\hat{W}_{t+1}$ denotes a random variable whose distribution coincides with the posterior predictive distribution of the wealth calculated at time point $t+1$.

In Theorem 8 we present the stochastic representations of the posterior predictive distribution of $\hat{W}_{t+1}$ with the proof given in the appendix. The symbol t_d stands for the standard univariate t-distribution with d degrees of freedom.

Theorem 8. *Under the assumption of Theorem 3 we get:*

(a) Under the diffuse prior (2.5), the stochastic representation of the posterior predictive distribution of W_{t+1} is given by

$$\begin{aligned} \widehat{W}_{t+1} \overset{d}{=} W_t\Bigg(&1 + r_{f,t+1} + \mathbf{v}_t^\top(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}) \\ &+ \sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,d}\mathbf{v}_t}\left(\frac{t_1}{\sqrt{n(n-k)}} + \sqrt{1+\frac{t_1^2}{n-k}}\frac{t_2}{\sqrt{n-k+1}}\right)\Bigg)\end{aligned}$$

where t_1 and t_2 are independent with $t_1 \sim t_{n-k}$ and $t_2 \sim t_{n-k+1}$.

(b) *Under the conjugate prior* (2.6) *and* (2.7), *the stochastic representation of the posterior predictive distribution of* W_{t+1} *is given by*

$$\begin{aligned}\widehat{W}_{t+1} &\stackrel{d}{=} W_t\Bigg(1 + r_{f,t+1} + \mathbf{v}_t^\top(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}) \\ &+ \sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,c}\mathbf{v}_t}\left(\frac{t_1}{\sqrt{(n+r_0)(n+d_0-2k)}} + \sqrt{1+\frac{t_1^2}{n+d_0-2k}}\frac{t_2}{\sqrt{n+d_0-2k+1}}\right)\Bigg),\end{aligned}$$

where t_1 *and* t_2 *are independent with* $t_1 \sim t_{n+d_0-2k}$ *and* $t_2 \sim t_{n+d_0-2k+1}$.

The results in Theorem 8 are very useful in analyzing the behavior of the investor's wealth during the whole investment period as well as at the final point T. It allows: (i) to calculate with which probability the investor can become bankrupt during the whole investment horizon at each time point; (ii) to construct the prediction intervals for the wealths at each time point of the investment period; (iii) to determine risk measures, like Value-at-Risk (VaR) and conditional VaR (CVaR), of the investment strategy during all times of the future reallocation; (iv) to specify a region where the final wealth belongs to with a high probability. We illustrate these results based on real data in Section 3.

2.2 Empirical study

2.2.1 Data description

The data used in the empirical study consist of weekly returns on twelve stocks from the FTSE 100, namely Barclays, Glaxo Smith Kline, Standard Life, Marks and Spencer, Burberry Group plc, HSBC, LLoyds Banking, NEXT plc, Rolls-Royce Holding, The Sage Group, Tesco plc and Unilever which represent a variety of branches with strong international activities. Since the parameters of the asset returns are not usually constant over a longer period of time, we disregard the use of monthly data which are closer to the normal distribution and choose weekly returns as a compromise between actuality and the assumption of conditional normality. As a risk-free rate we use the weekly returns on the three-months US treasury bill.

The portfolio weights are estimated using a rolling window estimation with different sample sizes of $n \in \{52, 78, 104, 130\}$ corresponding to one year up to two and a half years of weekly data in steps of six months. The portfolio runs from 6.6.2016 until 5.9.2016 ($T = 13$) covering a precarious market situation due to Great Britains referendum to leave the European Union on 23.06.2016. The gross returns of these assets are given in Figure 2.1. Especially Barclays suffered a loss of nearly 10 % in the week after the Brexit decision but also suffered losses in the

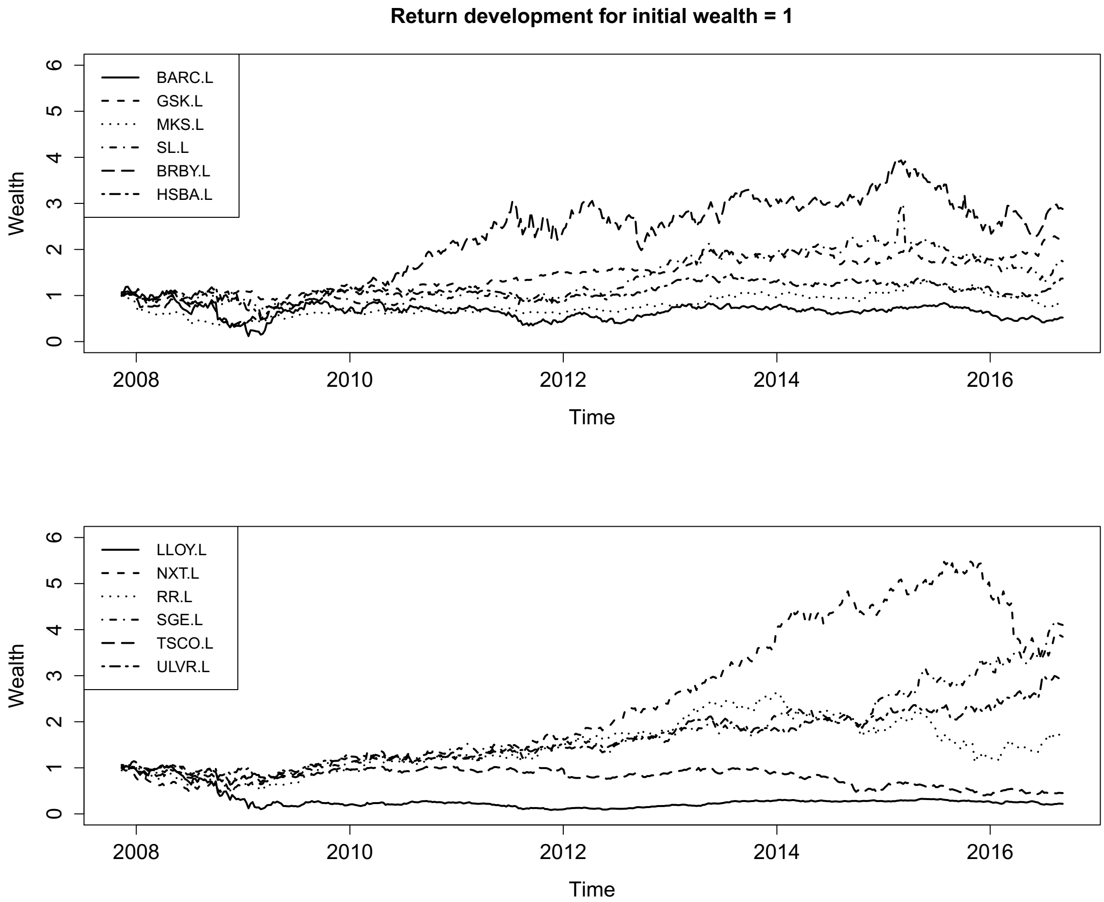

Figure 2.1: Development of the gross returns for the twelve assets considered in the portfolio.

weeks prior to the Brexit. HSBC announced that significant parts of her banking operations is moved from the City of London to different locations as a direct reaction to the referendum and it is rumoured that Lloyds seeks for a German banking licence as a consequence to the Brexit. The returns of the Marks and Spencer share were not as affected by the Brexit but the company reported that consumer confidence would be weakened in the days prior to the Brexit. This also implies price uncertainty for domestic consumer products due to a decline of the pound losing almost a fifth of his value against the dollar after the Brexit vote, which was emphasized for example by Tesco and Unilever. But Glaxo Smith Kline and Standard Life seem to be unaffected by the Brexit decision, yielding even positive returns. Rolls Royce, after all, faced significant losses in the beginning of 2016 and is hit by the Brexit vote severely, since they need to hedge a huge amount of British pounds against currency fluctuations because most of the contracts in aerospace are conducted in dollars.

2.2.2 Posterior distribution of the weights

Due to Theorem 4 it is possible to access the posterior distribution of the weights directly. The weights can be sampled using the following procedure:

1. Generate independently
   - $\eta \sim \chi^2_n$ under the diffuse prior or $\eta \sim \chi^2_{n+d_0-k}$ under the conjugate prior
   - $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$
   - $Q \sim \mathcal{F}(k, n-k)$ under the diffuse prior or $Q \sim \mathcal{F}(k, n+d_0-2k)$ under the conjugate prior
   - $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k) \rightarrowtail \mathbf{u} = \mathbf{Z}/\sqrt{\mathbf{Z}'\mathbf{Z}}$
2. Compute the vector of portfolio weights by using the stochastic representation (2.8) for the diffuse prior or (2.9) for the conjugate prior.
3. Repeat steps (1) and (2) B times.

The implementation of this simulation procedure leads to sequences of optimal portfolio weights of size B at each time point of the investment period, from which using their sample distribution we approximate the posterior distributions of the weights as well as their important quantiles from these distributions and the credible sets for portfolio weights. It is remarkable that all computations can easily be done by generating samples from the well known univariate distributions and high numerical precision could be achieved by choosing the corresponding value of B.

In Figures 2.2.2 and 2.2.2, we analyze the finite-sample behavior of the results presented in Theorem 7. Namely, we investigate the speed of convergence of the posterior distribution of the optimal portfolio weights to the corresponding asymptotic distribution which is a normal distribution according to Theorem 7 for both priors. The choice of the hyperparameters $\mathbf{m}_0$ and $\mathbf{S}_0$ in the case of the conjugate prior are of particular interest. According to the Bayesian paradigm, $\mathbf{m}_0$ and $\mathbf{S}_0$ represent the correct belief of the decision maker. In practice, however, there are several data driven methods how to replace $\mathbf{m}_0$ and $\mathbf{S}_0$ by data-dependent values $\hat{\mathbf{m}}_0$ and $\hat{\mathbf{S}}_0$. We make use of the empirical Bayes approach (see section 2.4 for the derivation of the formulas) which is applied to the weekly data of the returns on the corresponding assets directly from the time period before the empirical counterparts of the portfolio weights are estimated, always with the same time window. Namely, they are given by

$$\hat{\mathbf{m}}_0 \quad = \quad \overline{\mathbf{x}}_{n-t} \quad \text{and} \quad \hat{\mathbf{S}}_0 = \frac{(d_0-k-1)(n-1)}{n}\mathbf{S}_{n-t}$$

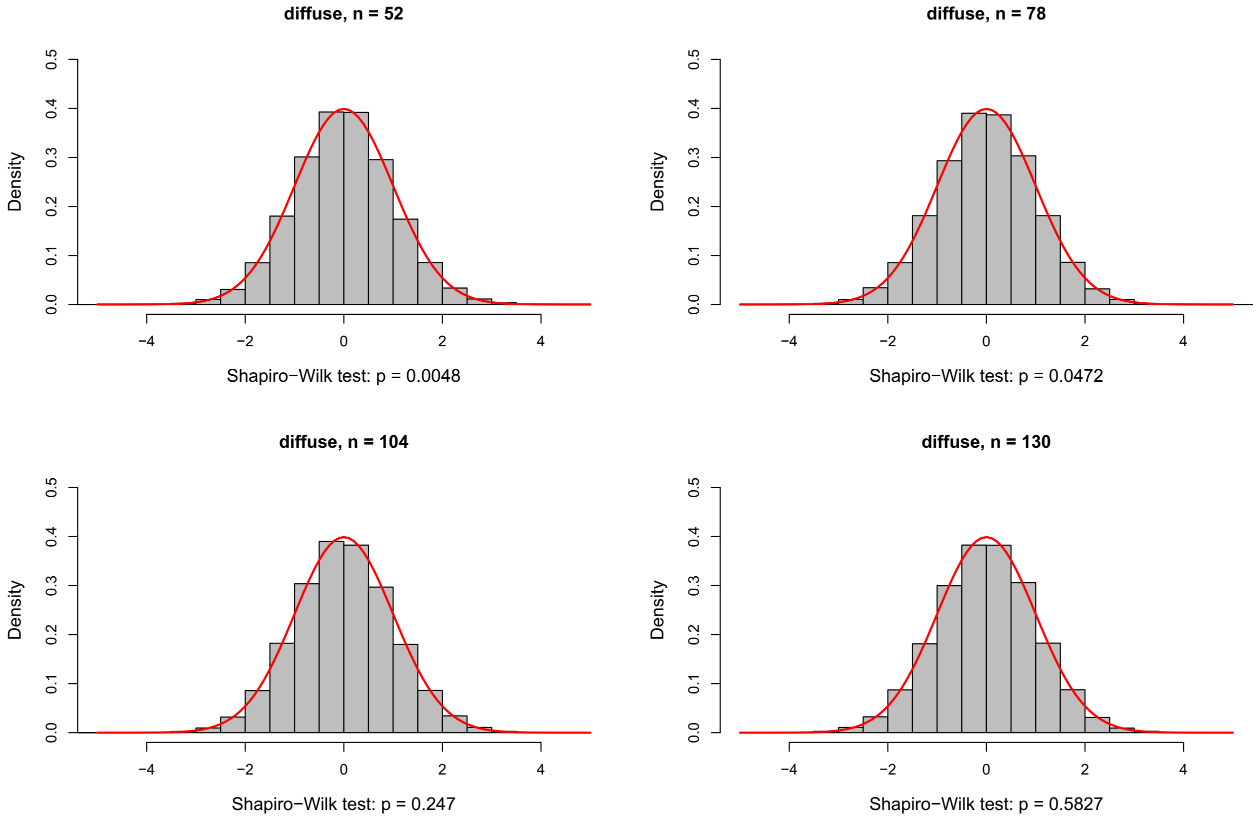

Figure 2.2: Histograms of the standardized Glaxo Smith Kline (GSK) weight for the diffuse prior.
The hypothesis that the weight is normally distributed can not be rejected for common significance levels when the sample size is larger than $n = 100$.

with the derivation moved to the appendix (Section 5.2). The prior parameters for $t > 1$ are estimated using a rolling window starting in the corresponding period. We set d_0 equal to the number of observations in the pres-sample period, i.e., $d_0 = n$.

We set $B = 10^5$ for draws from the stochastic representations of Theorem 4 and compare the standardized weight of Glaxo Smith Kline (GSK) calculated for the priod $T - 1$ in the case of several sample sizes $n \in \{52, 78, 104, 130\}$. The corresponding histograms are given in Figure 2.2.2 for the diffuse prior and in Figure 2.2.2 for the conjugate prior. In both figures we also present the p-values of the Shapiro-Wilk test, indicating if the standardized weights follow a standard normal distribution. This hypothesis is rejected for $n = 52$ and $n = 78$ in the case of the diffuse prior for a common significance level of 5 % but it cannot be rejected at this level for larger sample sizes. Stronger results are obtained in the case of the conjugate prior, where

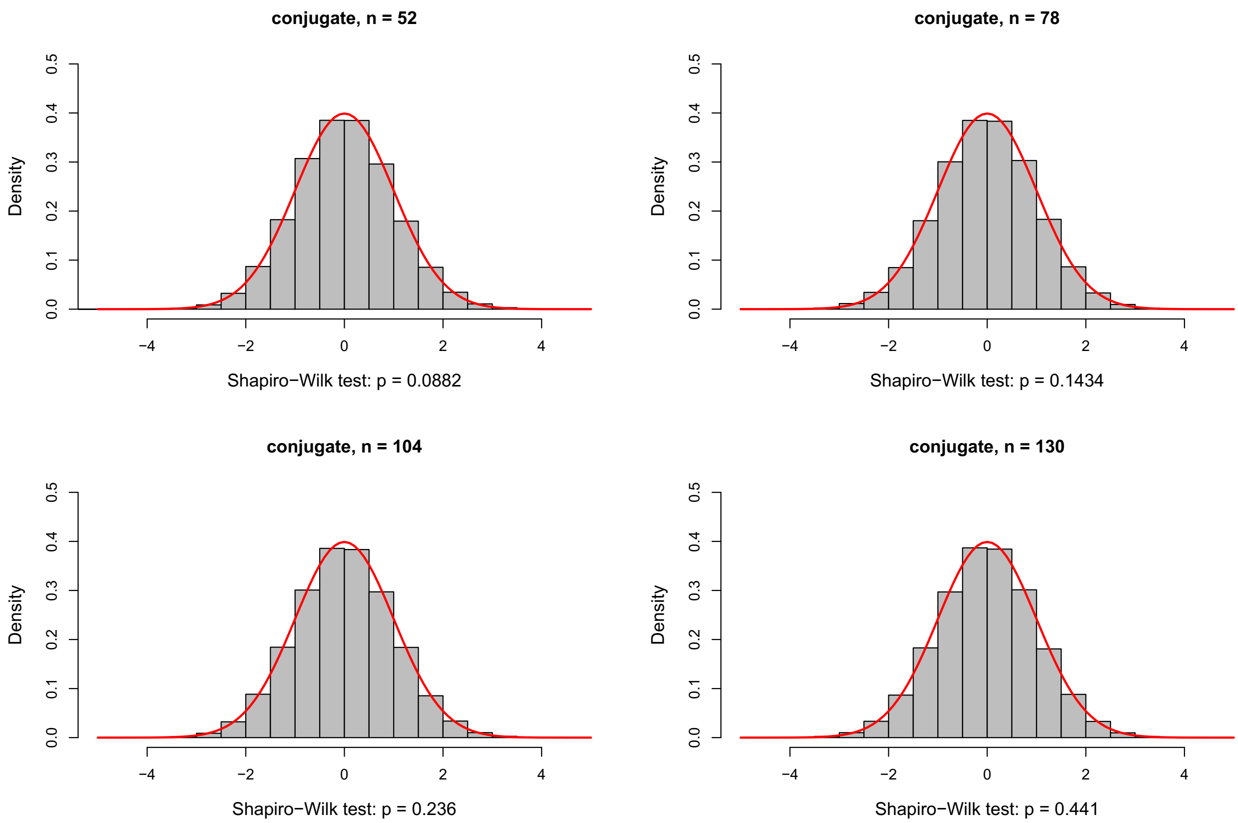

Figure 2.3: Histograms of the standardized Glaxo Smith Kline (GSK) weight for the conjugate prior.
The hypothesis that the weight is normally distributed can not be rejected for common significance levels in the case of all considered sample sizes.

the null hypothesis cannot be rejected at 5 % level for all considered sample sizes. We therefore conclude that the approximate distribution of Theorem 7 works reasonably well.

2.2.3 Wealth development and credibility intervals

Since the main purpose of investing is making money, investors are therefore interested in how much money they made during an investment period. We focus again on the same investment period covering the Brexit-referendum as in the previous subsection.

During the lifetime of the portfolio, no bankruptcy occurred. But more importantly, the stochastic representation for the posterior predictive distribution given in Theorem 6 can be used to calculate credible intervals for the wealth. By generating $B = 10^5$ draws from Theorem 6 and calculating the 95 % credible intervals, we generate upper and lower bounds for the wealth

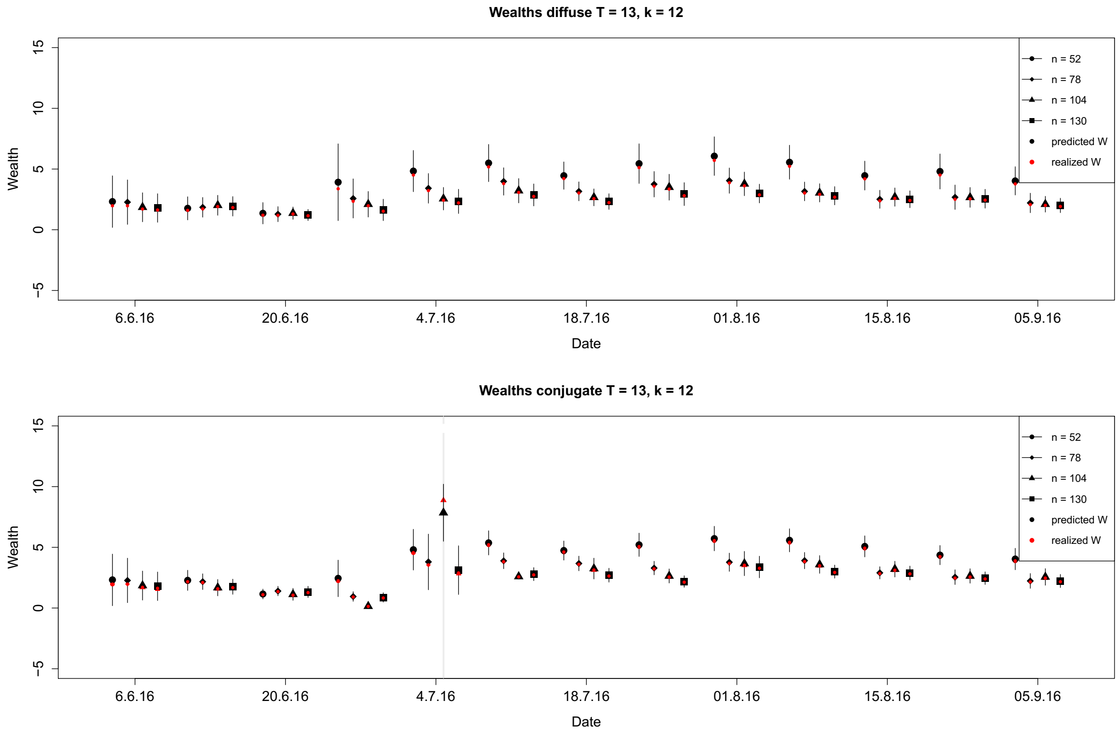

Figure 2.4: Wealth development and 95% credible intervals for the diffuse prior and for the conjugate prior.
The wealth for smaller n is almost always higher compared to a portfolio estimated with larger n, while the credible intervals are much narrower for larger n.

in the specific period. These intervals together with the predicted and realized wealths are shown in Figure 2.2.3. We observe a difference in the width of the intervals for lower and larger sample sizes which was expected. The credible intervals are considerably smaller for $n \in \{104, 130\}$ compared to smaller n. Note that the sample size has to be sufficiently large in relation to the number of assets. Otherwise, the credible intervals are inflated due to massive estimation uncertainty known as the curse of dimensionality.

It might happen that both the diffuse and the conjugate priors do not perform well when the sample size increases. The reason for the diffuse prior is that the empirical counterparts might not describe the portfolio running period well, indicating a trade-off between the actuality and stability of the parameters. This problem is amplified for the conjugate prior since the prior parameters are determined using even more distant data. While the data-driven approach to the conjugate prior is somewhat realistic, it is not completely in line with the Bayesian paradigm. When the expectations and therefore the choice of hyperparameters are closer to the return behaviour after the Brexit, the results could be improved. Although this is consistent with the Bayesian paradigm, such an approach is of course not entirely practical but not impractical: using appropriate forecasting methods, other data driven methods can be applicable as long as they yield a reliable point estimate. This subjective approach emphasizes the possibility as well as the necessity to resemble realistic future market behaviour in the prior parameterization and it is left for future research.

2.2.4 Default probability

Due to the accessability of the posterior predictive distribution, we can also calculate the default probability of our portfolio at each time point, defined as the event that our wealth becomes negative at this point in time. The predictive probability of default can easily be determined by calculating the amount of defaults in relation to all draws, in this case $B = 10^5$. The development of the defaults is given in Figure 2.5. Again, we find a pattern resembling the credible intervals of the posterior predictive distribution illustrated in the previous section with no surprises.

Starting with the diffuse prior, we observe a slightly increased default probability on 27.6.2016, the week after the Brexit referendum. With the conjugate prior, this default probability is lower in the same week. Again, the peak for $n = 130$ of the diffuse prior again resembles the trade-off between parameter stability and actuality, resulting here in a slightly increased default probability. The default probability for the conjugate prior is slightly increased in the following week compared to the diffuse prior, presumably due to parameters relying on a wider estimation window.

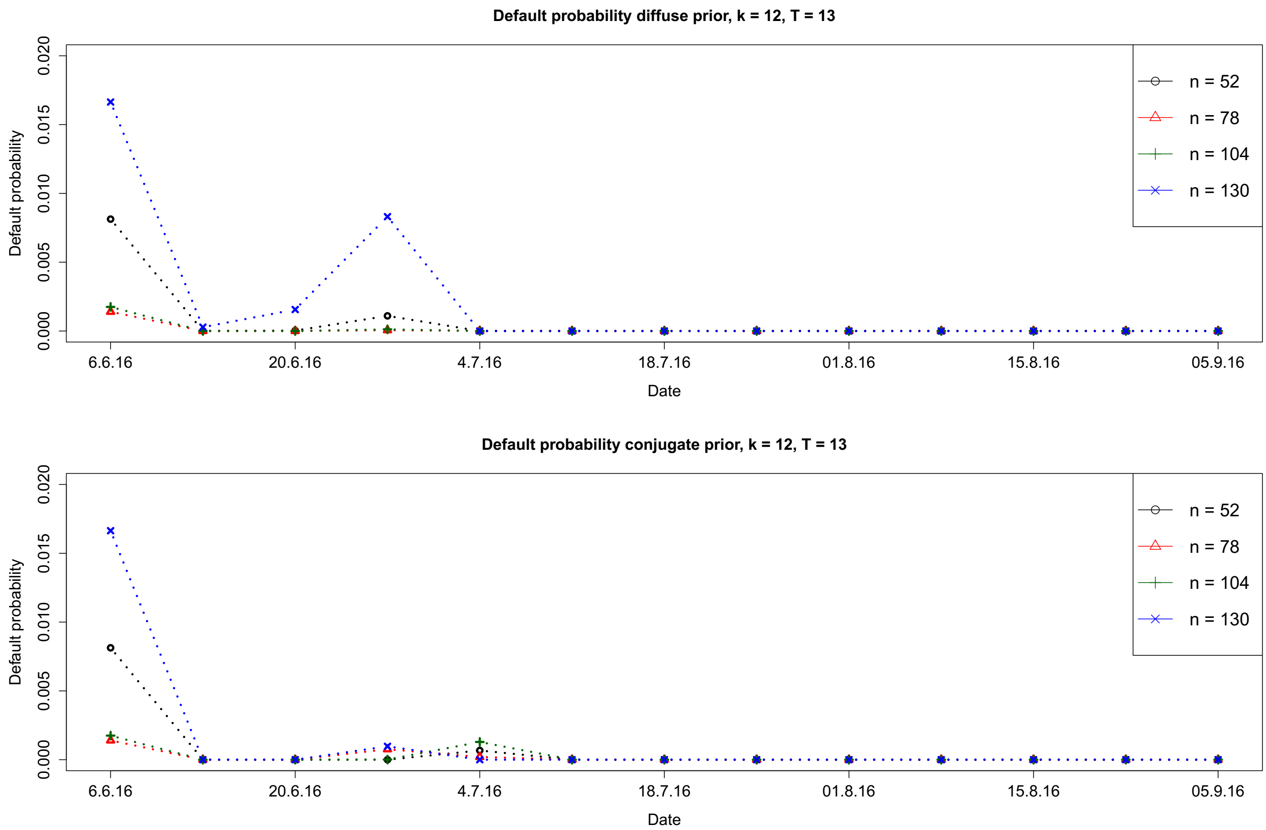

Figure 2.5: Default probabilities for the diffuse prior and for the conjugate prior.

2.3 Summary

In this chapter we consider the estimation of the multi-period portfolio for an exponential utility function in a Bayesian setting. Since the portfolio weights are given as the product of two multivariate/matrix-variate random quantities, accessing the distribution of the weights is a challenging task. By choosing the non-informative and the conjugate prior, the posterior distributions of the weights have pleasing properties since the conditional distribution of the precision matrix for a given return vector is an inverted Wishart distribution. With this insight we could use this well understood distribution (c.f. Muirhead (1982)) to derive stochastic representations for the weights which is a direct access to the posterior distribution. Furthermore, these representations also provide us with Bayesian estimates for the optimal portfolio weights together with their covariance matrix. In addition to this, we derive the posterior predictive distribution for the wealth which makes it possible to calculate the quantiles of the portfolio wealth at each time point of the investment period and it is therefore highly relevant for risk purposes. The method is then applied to real data from the FTSE 100 covering the period of the Brexit referendum. With these data we determine the posterior distribution of the weights, the predictive wealths in each period, the lower wealth quantiles as well as the default probability in every time period.

It turns out that the use of stochastic representations to generate the posterior distribution numerically is computationally highly efficient: the representations rely on samples from well known distributions and no MCMC methods are needed. In the empirical part of Section 3 it was demonstrated that these methods work well and are easy to implement. We have to emphasize several points: while the non-informative prior will yield results which coincide with the common frequentist case and is as easily to apply as the classical case, the conjugate or informative prior is said to involve a potentially large degree of subjectivity – sometimes implying that the frequentist approach or the non-informative prior would be objective. But we have to choose the sample size in all of these cases which is naturally a subjective choice with a huge effect on the performance of the portfolio as we demonstrate in Section 3. This trade-off between parameter actuality and parameter stability has to be faced by the practitioner. One advantage of the conjugate prior is of course that we can incorporate our beliefs regarding the future behaviour of the asset returns in our model which is not possible neither in the frequentist nor in the non-informative case. This is clearly at the core of every investment decision and reflects natural decision making. Nevertheless, the hyperparameters have to be chosen carefully and a rigorous sensitivity analysis is left for future research.

There are still other open research questions regarding the multi-period portfolio choice with exponential utility function which are left for future research. The present approach can be extended to the case with predictable variables as discussed in Bodnar et al. (2015b) in the case of the known parameters of the asset return distribution. This, however, is much more difficult due to the more complicated structure of the optimal portfolio weights and the dependence structure of the asset returns. Furthermore, the multi-period optimal portfolios obtained by using other utility functions can be estimated following the approach suggested in the chapter.

2.4 Proofs and Supplementary Material

In this part of the paper we present the proofs of the theoretical results. First, we note that the derived posterior distributions under the diffuse prior and under the conjugate prior in Proposition 2 have a similar structure. For that reason, we formulate and prove some lemmas from which the results in both cases of the diffuse prior and the conjugate prior follow.

Lemma 1. *Let*

$$\mathbf{\Omega}|\boldsymbol{\nu}, \mathbf{y} \sim \mathcal{IW}_k(k_y, \mathbf{S}_y^*(\boldsymbol{\nu})) \quad \text{and} \quad \boldsymbol{\nu}|\mathbf{y} \sim t_k\left(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y\right),$$

where $\mathbf{S}_y^*(\boldsymbol{\nu}) = v_y(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top)$ *and let* $\mathbf{M}$ *be a* $p \times k$*-dimensional matrix of constants. Then the stochastic representation of* $\mathbf{M}\mathbf{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a})$ *is given by*

$$\begin{aligned} &\mathbf{M}\mathbf{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a}) \stackrel{d}{=} \eta \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a}) \\ + \quad &\sqrt{\eta}\left((\boldsymbol{\nu} - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a}) \cdot \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top - \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu} - \mathbf{a})(\boldsymbol{\nu} - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top\right)^{1/2} \mathbf{z}_0, \end{aligned}$$

where $\eta \sim \chi^2_{k_y-k-1}$*,* $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$*, and* $\boldsymbol{\nu}|\mathbf{y} \sim t_k\left(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y\right)$*; moreover,* $\eta, \mathbf{z}_0$ *and* $\boldsymbol{\nu}$ *are mutually independent.*

Proof of Lemma 1. Since $\mathbf{\Omega}^* \stackrel{d}{=} \mathbf{\Omega}|\boldsymbol{\nu} = \boldsymbol{\nu}^*, \mathbf{y} \sim \mathcal{IW}_k(k_y, \mathbf{S}_y^*(\boldsymbol{\nu}^*))$ and, consequently, $\mathbf{\Omega}^{*\,-1} \sim \mathcal{W}_k(k_y - k - 1, \mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1})$ (c.f., Theorem 3.4.1 in Gupta and Nagar (2000), it holds that (see, e.g., Theorem 3.2.5 in Muirhead (1982))

$$\mathbf{\Xi}^* = \widetilde{\mathbf{M}}\mathbf{\Omega}^{*\,-1}\widetilde{\mathbf{M}}^\top \sim \mathcal{W}_k(k_y - k - 1, \mathbf{V}^*),$$

with $\widetilde{\mathbf{M}} = (\mathbf{M}^\top, \boldsymbol{\nu}^* - \mathbf{a})^\top$ and $\mathbf{V}^* = \tilde{\mathbf{M}}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\tilde{\mathbf{M}}^\top$. Next, we partition $\mathbf{\Xi}^*$ and $\mathbf{V}^*$ in the following way

$$\mathbf{\Xi}^* = \begin{pmatrix} \mathbf{\Xi}_{11}^* & \mathbf{\Xi}_{12}^* \\ \mathbf{\Xi}_{21}^* & \Xi_{22}^* \end{pmatrix} = \begin{pmatrix} \mathbf{M}\mathbf{\Omega}^{*\,-1}\mathbf{M}^\top & (\boldsymbol{\nu}^* - \mathbf{a})^\top \mathbf{\Omega}^{*\,-1}\mathbf{M}^\top \\ \mathbf{M}\mathbf{\Omega}^{*\,-1}(\boldsymbol{\nu}^* - \mathbf{a}) & (\boldsymbol{\nu}^* - \mathbf{a})^\top \mathbf{\Omega}^{*\,-1}(\boldsymbol{\nu}^* - \mathbf{a}) \end{pmatrix}$$

and

$$\mathbf{V}^* = \begin{pmatrix} \mathbf{V}_{11}^* & \mathbf{V}_{12}^* \\ \mathbf{V}_{21}^* & V_{22}^* \end{pmatrix} = \begin{pmatrix} \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\mathbf{M}^\top & (\boldsymbol{\nu}^* - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\mathbf{M}^\top \\ \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}(\boldsymbol{\nu}^* - \mathbf{a}) & (\boldsymbol{\nu}^* - \mathbf{a})^\top \mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}(\boldsymbol{\nu}^* - \mathbf{a}) \end{pmatrix}.$$

The application of Theorem 3.2.10 in Muirhead (1982) yields

$$\mathbf{\Xi}_{12}^*|\Xi_{22}^* \sim \mathcal{N}_p(\mathbf{V}_{12}V_{22}^{-1}\Xi_{22}^*, \mathbf{V}_{11\cdot 2}\Xi_{22}^*) \quad \text{with} \quad \mathbf{V}_{11\cdot 2} = \mathbf{V}_{11} - \frac{\mathbf{V}_{12}\mathbf{V}_{21}}{V_{22}}.$$

Defining $\eta = \Xi_{22}^*/V_{22}$ and using Theorem 3.2.8 of Muirhead (1982) we get that $\eta \sim \chi^2_{k_y-k-1}$. Since the $\chi^2_{k_y-k-1}$-distribution is independent of $\boldsymbol{\nu} = \boldsymbol{\nu}^*$ and $\mathbf{y}$ (on which the distribution of Ξ_{22}^* depends on by definition of $\mathbf{\Xi}^*$), it is also the unconditional distribution of η as well as η is independent of both $\boldsymbol{\nu}$ and $\mathbf{y}$. Thus, the stochastic representation of $\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu}-\mathbf{a})$ is given by

$$\begin{aligned}\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu}-\mathbf{a}) &\overset{d}{=} \eta\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu}-\mathbf{a})+\sqrt{\eta}\Big((\boldsymbol{\nu}-\mathbf{a})^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu}-\mathbf{a})\cdot\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top\\ &- \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}(\boldsymbol{\nu}-\mathbf{a})(\boldsymbol{\nu}-\mathbf{a})^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top\Big)^{1/2}\mathbf{z}_0,\end{aligned}$$

where $\eta\sim\chi^2_{k_y-k-1}$, $\mathbf{z}_0\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, and $\boldsymbol{\nu}|\mathbf{y}\sim t_k\left(d_y,\mathbf{m}_y,\mathbf{S}_y/d_y\right)$; moreover, η, $\mathbf{z}_0$ and $\boldsymbol{\nu}$ are mutually independent. This completes the proof of the lemma. □

Proof of Theorem 3. The results of Theorem 3 follow from Lemma 1 with $\mathbf{M}=C_t\mathbf{L}$, $\boldsymbol{\Sigma}=\boldsymbol{\Omega}$, $\boldsymbol{\nu}=\boldsymbol{\mu}$, $\mathbf{a}=r_{f,t+1}\mathbf{1}$ and

(a) $k_y=n+k+1$, $d_y=n-k$, $v_y=n$, $\mathbf{m}_y=\overline{\mathbf{x}}_{t,d}$, $\mathbf{S}_y=\mathbf{S}_{t,d}/n$, and $\mathbf{S}_y^*(\boldsymbol{\nu})=\mathbf{S}_{t,d}^*(\boldsymbol{\mu})$ in the case of the diffuse prior;

(b) $k_y=n+d_0+1$, $d_y=n+d_0-2k$, $v_y=n+r_0$, $\mathbf{m}_y=\overline{\mathbf{x}}_{t,c}$, $\mathbf{S}_y=\mathbf{S}_{t,c}/(n+r_0)$, and $\mathbf{S}_y^*(\boldsymbol{\nu})=\mathbf{S}_{t,c}^*(\boldsymbol{\mu})$ in the case of the conjugate prior.

□

Lemma 2. *Under the conditions of Lemma 1, we get the following stochastic representation of* $\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu}-\mathbf{a})$ *expressed as*

$$\mathbf{M}\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu}-\mathbf{a})\overset{d}{=}v_y^{-1}\eta\mathbf{M}\boldsymbol{\zeta}+v_y^{-1}\sqrt{\eta}\left(\epsilon\mathbf{M}\boldsymbol{\Upsilon}\mathbf{M}^\top-\mathbf{M}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0,$$

with

$$\begin{aligned}\epsilon &= \epsilon(Q,\mathbf{u})=(\mathbf{m}_y-\mathbf{a})^\top\mathbf{S}_y^{-1}(\mathbf{m}_y-\mathbf{a})+2\frac{\sqrt{kQ/d_y}}{1+kQ/d_y}(\mathbf{m}_y-\mathbf{a})^\top\mathbf{S}_y^{-1/2}\mathbf{u}\\ &+ \frac{kQ/d_y}{1+kQ/d_y}-\frac{kQ/d_y}{1+kQ/d_y}\left((\mathbf{m}_y-\mathbf{a})^\top\mathbf{S}_y^{-1/2}\mathbf{u}\right)^2,\\ \boldsymbol{\zeta} &= \boldsymbol{\zeta}(Q,\mathbf{u})=\mathbf{S}_y^{-1}(\mathbf{m}_y-\mathbf{a})+\frac{\sqrt{kQ/d_y}}{1+kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}-\frac{kQ/d_y}{1+kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top\mathbf{S}_y^{-1/2}(\mathbf{m}_y-\mathbf{a}),\\ \boldsymbol{\Upsilon} &= \boldsymbol{\Upsilon}(Q,\mathbf{u})=\mathbf{S}_y^{-1}-\frac{kQ/d_y}{1+kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top\mathbf{S}_y^{-1/2},\end{aligned}$$

where $\eta\sim\chi^2_{k_y-k-1}$, $\mathbf{z}_0\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $Q\sim\mathcal{F}(k,d_y)$, *and* $\mathbf{u}$ *uniformly distributed on the unit sphere in* R^k; *moreover,* η, $\mathbf{z}_0$, Q, *and* $\mathbf{u}$ *are mutually independent.*

Proof of Lemma 2. The application of the Sherman-Morrison formula (see, e.g., p.125 in Meyer (2000)) yields

$$(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top)^{-1} = \mathbf{S}_y^{-1} - \frac{\mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}}{1 + (\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)} \tag{2.11}$$

Let

$$\mathbf{u} = \frac{\mathbf{S}_y^{-1/2}(\boldsymbol{\nu} - \mathbf{m}_y)}{\sqrt{(\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)}} \quad \text{and} \quad Q = d_y(\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)/k. \tag{2.12}$$

Since $\boldsymbol{\nu}|\mathbf{y} \sim t_k(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y)$ and that the multivariate t-distribution belongs to the class of the elliptically contoured distributions, we obtain that $\mathbf{u}$ and Q are independent, and $\mathbf{u}$ is uniformly distributed on the unit sphere in R^k (see Theorem 2.15 of Gupta et al. (2013)). Moreover, from the properties of the multivariate t-distribution (see p. 19 of Kotz and Nadarajah (2004)), we get that $Q \sim \mathcal{F}(k, d_y)$, i.e., Q has an $\mathcal{F}$-distribution with k and d_y degrees of freedom.

Hence, the application of the (2.11) and (2.12) leads to

$$\begin{aligned}
&(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top)^{-1} = \mathbf{S}_y^{-1} - \frac{kQ/d_y}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_y^{-1/2}, \\
&(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top)^{-1}(\boldsymbol{\nu} - \mathbf{a}) \\
= \ & \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{a}) - \frac{\mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y + \mathbf{m}_y - \mathbf{a})}{1 + (\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)} \\
= \ & \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + \frac{\mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)}{1 + (\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)} - \frac{\mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})}{1 + (\boldsymbol{\nu} - \mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu} - \mathbf{m}_y)} \\
= \ & \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + \frac{\sqrt{kQ/d_y}}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u} - \frac{kQ/d_y}{1 + kQ/d_y}\mathbf{S}_y^{-1/2}\mathbf{u}\mathbf{u}^\top \mathbf{S}_y^{-1/2}(\mathbf{m}_y - \mathbf{a}),
\end{aligned}$$

and

$$\begin{aligned}
&(\boldsymbol{\nu} - \mathbf{a})^\top(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top)^{-1}(\boldsymbol{\nu} - \mathbf{a}) \\
= \ & (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) + 2\frac{(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u}\sqrt{kQ/d_y}}{1 + kQ/d_y} \\
+ \ & \frac{kQ/d_y}{1 + kQ/d_y} - \frac{kQ/d_y}{1 + kQ/d_y}\left((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u}\right)^2.
\end{aligned}$$

Putting the above results together we obtain the statement of the lemma. □

Proof of Theorem 4. The results of Theorem 4 are obtained by using Lemma 2 with $\mathbf{M} = C_t\mathbf{L}$,

$\mathbf{\Sigma} = \mathbf{\Omega}$, $\boldsymbol{\nu} = \boldsymbol{\mu}$, $\mathbf{a} = r_{f,t+1}\mathbf{1}$ and

(a) $k_y = n+k+1$, $d_y = n-k$, $v_y = n$, $\mathbf{m}_y - \mathbf{a} = \overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}$, $\mathbf{S}_y = \mathbf{S}_{t,d}/n$, and $\mathbf{S}^*_y(\boldsymbol{\nu}) = \mathbf{S}^*_{t,d}(\boldsymbol{\mu})$ in the case of the diffuse prior;

(b) $k_y = n + d_0 + 1$, $d_y = n + d_0 - 2k$, $v_y = n + r_0$, $\mathbf{m}_y - \mathbf{a} = \overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}$, $\mathbf{S}_y = \mathbf{S}_{t,c}/(n + r_0)$, and $\mathbf{S}^*_y(\boldsymbol{\nu}) = \mathbf{S}^*_{t,c}(\boldsymbol{\mu})$ in the case of the conjugate prior.

□

Proof of Theorem 5. The proof of the theorem is based on the stochastic representations obtained in Theorem 4. Let $\mathbf{l}$ be an arbitrary k-dimensional vector of constants.

(a) Using that η, $\mathbf{z}_0$ Q, and $\mathbf{u}$ are independent and that $\mathbb{E}(\mathbf{z}_0) = \mathbf{0}$, in the case of the diffuse prior we get

$$\mathbb{E}(\mathbf{l}^\top \boldsymbol{w}_t|\mathbf{x}_{t,n}) \quad = \quad C_t \mathbb{E}(\eta)\mathbf{l}^\top \mathbb{E}(\boldsymbol{\zeta}_d)$$

with $\mathbb{E}(\eta) = n$ and

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\zeta}_d|\mathbf{x}_{t,n}) \quad &= \quad \mathbf{S}^{-1}_{t,d}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) + \frac{1}{\sqrt{n}}\mathbb{E}\left(\frac{\sqrt{kQ/(n-k)}}{1 + kQ/(n-k)}\mathbf{S}^{-1/2}_{t,d}\right)\mathbb{E}(\mathbf{u}) \\
&- \quad \mathbb{E}\left(\frac{kQ/(n-k)}{1 + kQ/(n-k)}\mathbf{S}^{-1/2}_{t,d}\right)\mathbb{E}(\mathbf{u}\mathbf{u}^\top)\mathbf{S}^{-1/2}_{t,d}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) \\
&= \quad \mathbf{S}^{-1}_{t,d}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}) - \frac{k}{n}\frac{1}{k}\mathbf{S}^{-1}_{t,d}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1}),
\end{aligned}$$

where we use that $E(\mathbf{u}) = \mathbf{0}$ and $E(\mathbf{u}\mathbf{u}^T) = \frac{1}{k}\mathbf{I_k}$ (see, e.g. Gupta et al. (2013)) as well as the fact that if $Q \sim \mathcal{F}(k, n-k)$, then $\frac{k}{n-k}Q/\left(1 + \frac{k}{n-k}Q\right) \sim Beta\left(\frac{k}{2}, \frac{n-k}{2}\right)$. Hence,

$$E\left(\frac{\frac{k}{(n-k)}Q}{1 + \frac{k}{(n-k)}Q}\right) = \frac{k}{n}$$

and, consequently, since $\mathbf{l}$ was an arbitrary vector, we get

$$\mathbb{E}(\boldsymbol{w}_t|\mathbf{x}_{t,n}) \quad = \quad C_t(n-1)\mathbf{S}^{-1}_{t,d}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})\,.$$

(b) Similar computations as in part (a) leads to

$$\mathbb{E}(\boldsymbol{w}_t|\mathbf{x}_{t,n}) = C_t(n + d_0 - k - 1)\mathbf{S}^{-1}_{t,c}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$$

under the conjugate prior.

$\square$

Lemma 3. *Under the assumption of Lemma 2 with $\mathbf{M} = \mathbf{b}^\top : 1 \times k$, we get that*

$$
\begin{aligned}
& v_y^2 \mathbb{E}((\mathbf{b}^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a}))^2 | \mathbf{y}) = (k_y - k - 1)(k_y - k) \Bigg[\left(1 - \frac{2}{k + d_y} + \frac{2}{(k + d_y)(k + d_y + 2)} \right) c_{12}^2 \\
+ & \left(\frac{d_y}{(k + d_y)(k + d_y + 2)} + \frac{1}{(k + d_y)(k + d_y + 2)} c_2 \right) c_1 \Bigg] \\
+ & (k_y - k - 1) \Bigg[\left(\frac{k-1}{k + d_y} + \left(1 - \frac{1}{k} - \frac{1}{k + d_y} + \frac{1}{(k + d_y)(k + d_y + 2)} \right) c_2 \right) c_1 \\
+ & \frac{2}{(k + d_y)(k + d_y + 2)} c_{12}^2 \Bigg],
\end{aligned}
$$

where $c_1 = \mathbf{b}^\top \mathbf{S}_y^{-1} \mathbf{b}$, $c_2 = (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1} (\mathbf{m}_y - \mathbf{a})$, and $c_{12} = \mathbf{b}^\top \mathbf{S}_y^{-1} (\mathbf{m}_y - \mathbf{a})$.

Proof. The proof of the lemma is based on the stochastic representations from Lemma 2. Since η, $\mathbf{z}_0$, Q, and $\mathbf{u}$ are independent as well as $\mathbb{E}(\mathbf{z}_0) = \mathbf{0}$ and $\mathbb{E}(\mathbf{z}_0 \mathbf{z}_0^\top) = \mathbf{I}_p$, we obtain

$$
\begin{aligned}
& v_y^2 \mathbb{E}((\mathbf{b}^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{\nu} - \mathbf{a}))^2 | \mathbf{y}) = \mathbb{E}(\eta^2) \mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) + \mathbb{E}(\eta) \left(\mathbb{E}(\epsilon \mathbf{b}^\top \boldsymbol{\Upsilon} \mathbf{b} | \mathbf{y}) - \mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) \right) \\
= & (k_y - k - 1)(k_y - k) \mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) + (k_y - k - 1) \mathbb{E}(\epsilon \mathbf{b}^\top \boldsymbol{\Upsilon} \mathbf{b} | \mathbf{y})
\end{aligned}
$$

with $\mathbb{E}(\eta) = k_y - k - 1$ and $\mathbb{E}(\eta^2) = (k_y - k - 1)(k_y - k + 1)$.

The application of $E(\mathbf{u}\mathbf{u}^T) = \frac{1}{k} \mathbf{I}_k$ and the fact that all odd mixed moments of $\mathbf{u}$ are zero yield

$$
\begin{aligned}
\mathbb{E}((\mathbf{b}^\top \boldsymbol{\zeta})^2 | \mathbf{y}) &= (\mathbf{b}^\top \mathbf{S}_y^{-1} (\mathbf{m}_y - \mathbf{a}))^2 + \frac{1}{k} \mathbb{E} \left(\frac{kQ/d_y}{(1 + kQ/d_y)^2} \right) \mathbf{b}^\top \mathbf{S}_y^{-1} \mathbf{b} \\
&- \frac{2}{k} \mathbb{E} \left(\frac{kQ/d_y}{1 + kQ/d_y} \right) (\mathbf{b}^\top \mathbf{S}_y^{-1} (\mathbf{m}_y - \mathbf{a}))^2 \\
&+ E \left(\left(\frac{kQ/d_y}{1 + kQ/d_y} \right)^2 \right) E \left((\mathbf{b}^\top \mathbf{S}_y^{-1/2} \mathbf{U})^2 ((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2} \mathbf{U})^2 | \mathbf{y} \right)
\end{aligned}
$$

and

$$\begin{aligned}
\mathbb{E}\left(\epsilon \mathbf{b}^\top \boldsymbol{\Upsilon} \mathbf{b}|\mathbf{y}\right) &= (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} + \mathbb{E}\left(\frac{kQ/d_y}{1+kQ/d_y}\right)\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} \\
&- \frac{1}{k}\mathbb{E}\left(\frac{kQ/d_y}{1+kQ/d_y}\right)(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} \\
&- \frac{1}{k}(\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} - \frac{1}{k}\mathbb{E}\left(\frac{kQ/d_y}{1+kQ/d_y}\right)\mathbf{b}^T \mathbf{S}_y^{-1}\mathbf{b} \\
&+ E\left(\left(\frac{kQ/d_y}{1+kQ/d_y}\right)^2\right)\mathbb{E}\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2|\mathbf{y}\right).
\end{aligned}$$

Since $\frac{kQ/d_y}{1+kQ/d_y}$ has a beta distribution with $k/2$ and $d_y/2$ degrees of freedom, we obtain

$$\begin{aligned}
E\left(\frac{kQ/d_y}{1+kQ/d_y}\right) &= \frac{k}{k+d_y}, \\
E\left(\frac{kQ/d_y}{1+kQ/d_y}\right)^2 &= \frac{2kd_y + k^2(k+d_y+2)}{(k+d_y)^2(k+d_y+2)} = \frac{k(k+2)}{(k+d_y)(k+d_y+2)}.
\end{aligned}$$

Furthermore, using $Q \sim \mathcal{F}(k, d_y)$, we get

$$\begin{aligned}
E\left[\frac{kQ/d_y}{(1+kQ/d_y)^2}\right] &= \frac{1}{n_0}\int_0^\infty \frac{kt/d_y}{(1+kt/d_y)^2}\frac{1}{B\left(\frac{k}{2},\frac{d_y}{2}\right)}\left(\frac{k}{d_y}\right)^{k/2} t^{k/2-1}\left(1+\frac{k}{d_y}t\right)^{-(k+d_y)/2} dt \\
&= \frac{1}{B\left(\frac{k}{2},\frac{d_y}{2}\right)}\int_0^\infty \left(\frac{k}{d_y}\right)^{(k+2)/2} t^{(k+2)/2-1}\left(1+\frac{k}{d_y}t\right)^{-(k+d_y+4)/2} dt \\
&= \frac{B\left(\frac{k+2}{2},\frac{d_y+2}{2}\right)}{B\left(\frac{k}{2},\frac{d_y}{2}\right)} = \frac{kd_y}{(k+d_y)(k+d_y+2)},
\end{aligned}$$

where $B(\cdot,\cdot)$ stands for the beta function (see, Mathai and Provost (1992, p. 256)).

Next, we compute $\mathbb{E}\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2((\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2|\mathbf{y}\right)$. Let $Q_N \sim \chi^2_k$ be independent of $\mathbf{u}$. Then $\sqrt{Q_N}\mathbf{u}$ has a multivariate standard normal distribution, i.e.

$$\begin{aligned}
\begin{pmatrix} \mathbf{b}^\top \mathbf{S}_y^{-1/2} \\ (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1/2}\end{pmatrix}\sqrt{Q_N}\mathbf{u} &\sim \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} \mathbf{b}^\top \mathbf{S}_y^{-1}\mathbf{b} & \mathbf{b}^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a}) \\ (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}\mathbf{b} & (\mathbf{m}_y - \mathbf{a})^\top \mathbf{S}_y^{-1}(\mathbf{m}_y - \mathbf{a})\end{pmatrix}\right) \\
&= \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} c_1 & c_{12} \\ c_{12} & c_2 \end{pmatrix}\right),
\end{aligned}$$

where c_1, c_2, and c_{12} are defined in the statement of Lemma 3. Hence,

$$\begin{aligned}
& \mathbb{E}\left((\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2((\mathbf{m}_y-\mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u})^2|\mathbf{y}\right) \\
= & \mathbb{E}\left[\left(\mathbf{b}^\top \mathbf{S}_y^{-1/2}\mathbf{u}\right)^2\left((\mathbf{m}_y-\mathbf{a})^\top \mathbf{S}_y^{-1/2}\mathbf{u}\right)^2|\mathbf{y}\right]\frac{\mathbb{E}(Q_N^2)}{\mathbb{E}(Q_N^2)} \\
= & \frac{\mathbb{E}\left[\left(\mathbf{b}^\top \mathbf{S}_y^{-1/2}\sqrt{Q_N}\mathbf{u}\right)^2\left((\mathbf{m}_y-\mathbf{a})^\top \sqrt{Q_N}\mathbf{S}_y^{-1/2}\mathbf{u}\right)^2|\mathbf{y}\right]}{\mathbb{E}(Q_N^2)} = \frac{c_1c_2+2c_{12}^2}{k(k+2)},
\end{aligned}$$

where the last equality follows from the Isserlis' theorem (c.f., Isserlis (1918)).

Hence,

$$\begin{aligned}
E(\mathbf{b}^\top\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{b}) &= c_{12}^2+\frac{1}{k}\frac{kd_y}{(k+d_y)(k+d_y+2)}c_1 \\
&- \frac{2}{k}\frac{k}{k+d_y}c_{12}^2+\frac{k(k+2)}{(k+d_y)(k+d_y+2)}\frac{c_1c_2+2c_{12}^2}{k(k+2)} \\
&= \left(1-\frac{2}{k+d_y}+\frac{2}{(k+d_y)(k+d_y+2)}\right)c_{12}^2 \\
&+ \left(\frac{d_y}{(k+d_y)(k+d_y+2)}+\frac{1}{(k+d_y)(k+d_y+2)}c_2\right)c_1
\end{aligned}$$

and

$$\begin{aligned}
& E\left(\epsilon\mathbf{b}^\top\boldsymbol{\Upsilon}\mathbf{b}\right) = c_1c_2+\frac{k}{k+d_y}c_1-\frac{1}{k}\frac{k}{k+d_y}c_1c_2 \\
- & \frac{1}{k}c_1c_2-\frac{1}{k}\frac{k}{k+d_y}c_1+\frac{k(k+2)}{(k+d_y)(k+d_y+2)}\frac{c_1c_2+2c_{12}^2}{k(k+2)} \\
= & \frac{2}{(k+d_y)(k+d_y+2)}c_{12}^2+\left(\frac{k-1}{(k+d_y)}+\left(1-\frac{1}{k}-\frac{1}{k+d_y}+\frac{1}{(k+d_y)(k+d_y+2)}\right)c_2\right)c_1\,.
\end{aligned}$$

□

Proof of Theorem 6. The results of Theorem 6 are obtained by using Lemma 3 with $\mathbf{b}=C_t\mathbf{l}$, $\boldsymbol{\Sigma}=\boldsymbol{\Omega}$, $\boldsymbol{\nu}=\boldsymbol{\mu}$, $\mathbf{a}=r_{f,t+1}\mathbf{1}$ and Theorem 5.

(a) In the case of the diffuse prior, using $k_y=n+k+1$, $d_y=n-k$, $v_y=n$, $\mathbf{m}_y-\mathbf{a}=\overline{\mathbf{x}}_{t,d}-r_{f,t+1}\mathbf{1}$, $\mathbf{S}_y=\mathbf{S}_{t,d}/n$, $c_1=nC_t^2\mathbf{l}^\top\mathbf{S}_{t,d}^{-1}\mathbf{l}$, $c_2=n(\overline{\mathbf{x}}_{t,d}-r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d}-r_{f,t+1}\mathbf{1})$, and

$c_{12} = nC_t\mathbf{l}^\top\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$ we get

$$
\begin{aligned}
&\mathbb{V}ar(\mathbf{l}^\top\boldsymbol{w}_t|\mathbf{y})\\
= \;& \frac{1}{n^2}\Bigg\{n(n+1)\left[\left(1-\frac{2}{n}+\frac{2}{n(n+2)}\right)c_{12}^2+\left(\frac{n-k}{n(n+2)}+\frac{1}{n(n+2)}c_2\right)c_1\right]\\
+ \;& n\left[\left(\frac{k-1}{n}+\left(1-\frac{1}{k}-\frac{1}{n}+\frac{1}{n(n+2)}\right)c_2\right)c_1+\frac{2}{n(n+2)}c_{12}^2\right]-(n-1)^2c_{12}^2\Bigg\}\\
= \;& \frac{n-1}{n^2}c_{12}^2+c_1\frac{1}{n^2}\left(\frac{n^2+k-2}{n+2}+\frac{n(k-1)}{k}c_2\right)\\
= \;& \mathbf{l}^\top\left(C_t^2\left((n-1)\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d}-r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,d}-r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1}+\left(\frac{n^2+k-2}{n(n+2)}+\frac{k-1}{k}b_d\right)\mathbf{S}_{t,d}^{-1}\right)\right)\mathbf{l}
\end{aligned}
$$

where $b_d = n(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$. Since $\mathbf{l}$ is an arbitrary vector, the results in part (a) follow.

(b) In the case of the conjugate prior, the application of $k_y = n + d_0 + 1$, $d_y = n + d_0 - 2k$, $v_y = n + r_0$, $\mathbf{m}_y - \mathbf{a} = \overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1}$, and $\mathbf{S}_y = \mathbf{S}_{t,c}/(n + r_0)$, $c_1 = (n + r_0)C_t^2\mathbf{l}^\top\mathbf{S}_{t,c}^{-1}\mathbf{l}$, $c_2 = (n+r_0)(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})$, and $c_{12} = (n+r_0)C_t\mathbf{l}^\top\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c} - r_{f,t+1}\mathbf{1})$.

leads to

$$
\begin{aligned}
& \mathbb{V}ar(\mathbf{l}^\top \boldsymbol{w}_t|\mathbf{y}) \\
= & \frac{1}{(n+r_0)^2}\Bigg\{(n+d_0-k)(n+d_0-k+1) \\
\times & \Bigg[\left(1-\frac{2}{n+d_0-k}+\frac{2}{(n+d_0-k)(n+d_0-k+2)}\right)c_{12}^2 \\
+ & \left(\frac{n+d_0-2k}{(n+d_0-k)(n+d_0-k+2)}+\frac{1}{(n+d_0-k)(n+d_0-k+2)}c_2\right)c_1\Bigg] \\
+ & (n+d_0-k)\Bigg[\left(\frac{k-1}{n+d_0-k}+\left(1-\frac{1}{k}-\frac{1}{n+d_0-k}+\frac{1}{(n+d_0-k)(n+d_0-k+2)}\right)c_2\right)c_1 \\
+ & \frac{2}{(n+d_0-k)(n+d_0-k+2)}c_{12}^2\Bigg]-(n+d_0-k-1)^2c_{12}^2\Bigg\} \\
= & \frac{1}{(n+r_0)^2}\left[\frac{n+d_0-k-1}{(n+d_0-k)^2}c_{12}^2+c_1\left(\frac{(n+d_0-k)^2+k-2}{n+d_0-k+2}+\frac{(n+d_0-k)(k-1)}{k}c_2\right)\right] \\
= & \mathbf{l}^\top\Bigg\{C_t^2\Bigg[(n+d_0-k-1)\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c}-r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,c}-r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1} \\
+ & \left(\frac{(n+d_0-k)^2+k-2}{(n+r_0)(n+d_0-k+2)}+\frac{(n+d_0-k)(k-1)}{(n+r_0)k}b_c\right)\mathbf{S}_{t,c}^{-1}\Bigg]\Bigg\}\mathbf{l}
\end{aligned}
$$

where $b_c=(n+r_0)(\overline{\mathbf{x}}_{t,c}-r_{f,t+1}\mathbf{1})^\top\mathbf{S}_{t,c}^{-1}(\overline{\mathbf{x}}_{t,c}-r_{f,t+1}\mathbf{1})$. Since $\mathbf{l}$ is an arbitrary vector, we get the statement of Theorem 6.(b).

□

Proof of Theorem 7. Let $\mathbf{l}$ be an arbitrary k-dimensional vector. From Theorem 3 with $\mathbf{L}=\mathbf{l}^\top$, we get the following stochastic representations of $\mathbf{L}\boldsymbol{w}_t$ under the diffuse prior and the conjugate prior expressed as

$$
\begin{aligned}
\mathbf{l}^\top\boldsymbol{w}_t \overset{d}{=} & C_t\eta\mathbf{l}^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu}-r_{f,t+1})+C_t\sqrt{\eta}\Big((\boldsymbol{\mu}-r_{f,t+1})^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu}-r_{f,t+1})\cdot\mathbf{l}^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}\mathbf{l} \\
- & \mathbf{l}^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu}-r_{f,t+1})(\boldsymbol{\mu}-r_{f,t+1})^\top\mathbf{S}_{t,d}^*(\boldsymbol{\mu})^{-1}\mathbf{l}\Big)^{1/2}\mathbf{z}_0,
\end{aligned}
$$

where $\eta \sim \chi^2_n$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n-k, \overline{\mathbf{x}}_{t,d}, \mathbf{S}_{t,d}/(n(n-k))\right)$, and

$$\begin{aligned}
\mathbf{l}^\top \boldsymbol{w}_t &\stackrel{d}{=} C_t \eta \mathbf{l}^\top \mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) \\
&+ C_t \sqrt{\eta}\Big((\boldsymbol{\mu} - r_{f,t+1})^\top \mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1}) \cdot \mathbf{l}^\top \mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}\mathbf{l} \\
&- \mathbf{l}^\top \mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}(\boldsymbol{\mu} - r_{f,t+1})(\boldsymbol{\mu} - r_{f,t+1})^\top \mathbf{S}^*_{t,c}(\boldsymbol{\mu})^{-1}\mathbf{l}\Big)^{1/2} \mathbf{z}_0,
\end{aligned}$$

where $\eta \sim \chi^2_{n+d_0-k}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$,
and $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n+d_0-2k, \overline{\mathbf{x}}_{t,c}, \mathbf{S}_{t,c}/((n+r_0)(n+d_0-2k))\right)$.

Moreover, since

$$\sqrt{n}\left(\begin{pmatrix} \eta/n \\ \mathbf{z}_0/\sqrt{n} \\ \boldsymbol{\mu} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{0} \\ \overline{\mathbf{x}}_{t,d} \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \breve{\mathbf{S}}_t \end{pmatrix}\right)$$

and

$$\sqrt{n}\left(\begin{pmatrix} \eta/n \\ \mathbf{z}_0/\sqrt{n} \\ \boldsymbol{\mu} \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{0} \\ \overline{\mathbf{x}}_{t,c} \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \breve{\mathbf{S}}_t \end{pmatrix}\right)$$

as $n \longrightarrow \infty$ as well as

$$\lim_{n\longrightarrow\infty} \overline{\mathbf{x}}_{t,c} = \breve{\mathbf{x}}_t = \lim_{n\longrightarrow\infty} \overline{\mathbf{x}}_{t,d}$$

and

$$\lim_{n\longrightarrow\infty} \frac{\mathbf{S}_{t,c}}{n+r_0} = \breve{\mathbf{S}}_t = \lim_{n\longrightarrow\infty} \frac{\mathbf{S}_{t,d}}{n-1},$$

the application of the delta method (c.f., (DasGupta, 2008, Theorem 3.7)) proves that

$$\sqrt{n}(\mathbf{l}^\top \boldsymbol{w}_t - \mathbf{l}^\top \hat{\boldsymbol{w}}_t)|\mathbf{x}_{t,n} \xrightarrow{d.} \mathcal{N}_k(\mathbf{0}, f_d)$$

and

$$\sqrt{n}(\mathbf{l}^\top \boldsymbol{w}_t - \mathbf{l}^\top \hat{\boldsymbol{w}}_t)|\mathbf{x}_{t,n} \xrightarrow{d.} \mathcal{N}_k(\mathbf{0}, f_c),$$

as $n \longrightarrow \infty$ under the diffuse prior and the conjugate prior, respectively.

Finally, the results of Theorem 6 yield

$$
\begin{aligned}
f_d &= \lim_{n\longrightarrow\infty} \mathbb{V}ar(\sqrt{n}\mathbf{l}^\top \boldsymbol{w}_t) = \lim_{n\longrightarrow\infty} \mathbf{l}^\top \Bigg\{ C_t^2 \Bigg(n(n-1)\mathbf{S}_{t,d}^{-1}(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})(\overline{\mathbf{x}}_{t,d} - r_{f,t+1}\mathbf{1})^\top \mathbf{S}_{t,d}^{-1} \\
&+ \left(\frac{n^2+k-2}{n(n+2)} + \frac{k-1}{k} b_d \right) \mathbf{S}_{t,d}^{-1} \Bigg) \Bigg\} \mathbf{l} \\
&= \mathbf{l}^\top \Bigg\{ C_t^2 \Bigg[\breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top \breve{\mathbf{S}}_t^{-1} \\
&+ \left(1 + \frac{k-1}{k}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top \breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1}) \right) \breve{\mathbf{S}}_t^{-1} \Bigg] \Bigg\} \mathbf{l}
\end{aligned}
$$

and, similarly,

$$
\begin{aligned}
f_c &= \mathbf{l}^\top \Bigg\{ C_t^2 \Bigg[\breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top \breve{\mathbf{S}}_t^{-1} \\
&+ \left(1 + \frac{k-1}{k}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1})^\top \breve{\mathbf{S}}_t^{-1}(\breve{\mathbf{x}}_t - r_{f,t+1}\mathbf{1}) \right) \breve{\mathbf{S}}_t^{-1} \Bigg] \Bigg\} \mathbf{l} = f_d.
\end{aligned}
$$

Since, for each $\mathbf{l}$ the linear combination $\mathbf{l}^\top \boldsymbol{w}_t$ is asymptotically normally distributed, then we also get that the vector of weights $\boldsymbol{w}_t$ is asymptotically normal. □

Proof of Theorem 8. Since $\mathbf{x}_{t+1}|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and it is conditionally independent of $\mathbf{x}_{t,n}$, we get

$$
\widehat{W}_{t+1}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{x}_{t,n} \sim \mathcal{N}(W_t(1 + r_{f,t+1} + \mathbf{v}_t^\top(\boldsymbol{\mu} - r_{f,t+1})), W_t^2 \mathbf{v}_t^\top \boldsymbol{\Sigma} \mathbf{v}_t).
$$

(a) In the case of the diffuse prior, we observe that

$$
\frac{\mathbf{v}_t^\top \boldsymbol{\Sigma} \mathbf{v}_t}{\mathbf{v}_t^\top \mathbf{S}_{t,d}(\boldsymbol{\mu})^* \mathbf{v}_t} \overset{d}{=} \frac{1}{\xi}, \tag{2.13}
$$

where $\xi \sim \chi^2_{n-k+1}$ and is independent of $\boldsymbol{\mu}$ (see, e.g., Theorem 3.2.13 in Muirhead (1982)). Then the stochastic representation of $\widehat{W}_{t+1}$ is given by

$$
\widehat{W}_{t+1} \overset{d}{=} W_t \left(1 + r_{f,t+1} + \mathbf{v}_t^\top(\boldsymbol{\mu} - r_{f,t+1}) + \frac{\sqrt{\mathbf{v}_t^\top \mathbf{S}_{t,d}(\boldsymbol{\mu})^* \mathbf{v}_t}}{\sqrt{n-k+1}} t_2 \right),
$$

where $t_2 \sim t_1(n-k+1, 0, 1)$ is independent of $\boldsymbol{\mu}$. Finally, from the properties of the multivariate t-distribution, we obtain

$$\mathbf{v}_t^\top(\boldsymbol{\mu}-\overline{\mathbf{x}}_{t,d}) \sim t_1\left(n-k,0,\frac{\mathbf{v}_t^\top\mathbf{S}_{t,d}\mathbf{v}_t}{n(n-k)}\right),$$

which leads to

$$\begin{aligned}\widehat{W}_{t+1} &\stackrel{d}{=} W_t\Bigg(1+r_{f,t+1}+\mathbf{v}_t^\top(\overline{\mathbf{x}}_{t,d}-r_{f,t+1})\\ &+ \sqrt{\mathbf{v}_t^\top\mathbf{S}_{t,d}\mathbf{v}_t}\left(\frac{t_1}{\sqrt{n(n-k)}}+\sqrt{1+\frac{t_1^2}{n-k}}\frac{t_2}{\sqrt{n-k+1}}\right)\Bigg),\end{aligned}$$

where t_1 and t_2 are independent with $t_1 \sim t_{n-k}$ and $t_2 \sim t_{n-k+1}$.

(b) Similarly, for the conjugate prior, it holds that

$$\frac{\mathbf{v}_t^\top\boldsymbol{\Sigma}\mathbf{v}_t}{\mathbf{v}_t^\top\mathbf{S}_{t,c}(\boldsymbol{\mu})^*\mathbf{v}_t} \stackrel{d}{=} \frac{1}{\xi}, \tag{2.14}$$

where $\xi \sim \chi^2_{n+d_0-2k+1}$ and is independent of $\boldsymbol{\mu}$. Then the stochastic representation of $\widehat{W}_{t+1}$ is given by

$$\widehat{W}_{t+1} \stackrel{d}{=} W_t\left(1+r_{f,t+1}+\mathbf{v}_t^\top(\boldsymbol{\mu}-r_{f,t+1})+\frac{\sqrt{\mathbf{v}_t^\top\mathbf{S}_{t,c}(\boldsymbol{\mu})^*\mathbf{v}_t}}{\sqrt{n+d_0-2k+1}}t_2\right),$$

where $t_2 \sim t_{n+d_0-2k+1}$ is independent of $\boldsymbol{\mu}$. From the properties of the multivariate t-distribution, we get

$$\mathbf{v}_t^\top(\boldsymbol{\mu}-\overline{\mathbf{x}}_{t,c}) \sim t_1\left(n+d_0-2k,0,\frac{\mathbf{v}_t^\top\mathbf{S}_{t,c}\mathbf{v}_t}{(n+r_0)(n+d_0-2k)}\right),$$

which leads to

$$\begin{aligned}\widehat{W}_{t+1} &\stackrel{d}{=} W_t\Bigg(1+r_{f,t+1}+\mathbf{v}_t^\top(\overline{\mathbf{x}}_{t,c}-r_{f,t+1})\\ &+ \sqrt{\mathbf{v}_t^\top\mathbf{S}_{t,c}\mathbf{v}_t}\left(\frac{t_1}{\sqrt{(n+r_0)(n+d_0-2k)}}+\sqrt{1+\frac{t_1^2}{n+d_0-2k}}\frac{t_2}{\sqrt{n+d_0-2k+1}}\right)\Bigg),\end{aligned}$$

where t_1 and t_2 are independent with $t_1 \sim t_{n+d_0-2k}$ and $t_2 \sim t_{n+d_0-2k+1}$.

□

Now we derive the empirical Bayes estimates for the hyperparameters of the conjugate prior $\mathbf{m}_0$ and $\mathbf{S}_0$. Given the sample $\mathbf{x}_{\tau,n}$ the empirical Bayes estimates for $\mathbf{m}_0$ and $\mathbf{S}_0$ are obtained by maximizing (see, e.g., Carlin and Louis (2000))

$$g(\mathbf{m}_0, \mathbf{S}_0) = \int_{\boldsymbol{\mu}} \int_{\boldsymbol{\Sigma}} L(\mathbf{x}_{t,n}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{d}\boldsymbol{\Sigma}\mathrm{d}\boldsymbol{\mu} \tag{2.15}$$

with respect to $\mathbf{m}_0$ and $\mathbf{S}_0$.

First, we calculate the integral in (2.15), ignoring the terms which do not depend on $\mathbf{m}_0$ and $\mathbf{S}_0$, to get

$$\begin{aligned}
g(\mathbf{m}_0, \mathbf{S}_0) &\propto \int_{\boldsymbol{\mu}} \int_{\boldsymbol{\Sigma}} L(\mathbf{x}_{t,n}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{d}\boldsymbol{\Sigma}\mathrm{d}\boldsymbol{\mu} \\
&\propto \int_{\boldsymbol{\mu}} \int_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{n}{2}(\bar{\mathbf{x}}_\tau - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}_\tau - \boldsymbol{\mu}) - \frac{n-1}{2} tr(\mathbf{S}_\tau \boldsymbol{\Sigma}^{-1})\right\} \\
&\times |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{r_0}{2}(\boldsymbol{\mu} - \mathbf{m}_0)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \mathbf{m}_0) -\right\} \\
&\times |\boldsymbol{\Sigma}|^{-d_0/2} |\mathbf{S}_0|^{(d_0-k-1)/2} \exp\left\{-\frac{1}{2} tr(\mathbf{S}_0 \boldsymbol{\Sigma}^{-1})\right\} \mathrm{d}\boldsymbol{\Sigma}\mathrm{d}\boldsymbol{\mu} \\
&= |\mathbf{S}_0|^{(d_0-k-1)/2} \int_{\boldsymbol{\mu}} \int_{\boldsymbol{\Sigma}} |\boldsymbol{\Sigma}|^{-(n+d_0+1)/2} \exp\left\{-\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{-1}\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0)\right)\right\} \mathrm{d}\boldsymbol{\Sigma}\mathrm{d}\boldsymbol{\mu} \\
&\propto |\mathbf{S}_0|^{(d_0-k-1)/2} \int_{\boldsymbol{\mu}} |\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0)|^{-(n+d_0-k)/2} \mathrm{d}\boldsymbol{\mu}\,,
\end{aligned}$$

where the last identity is obtained by recognizing that under the integral with respect to $\boldsymbol{\Sigma}$ we have a kernel of the density function of $\mathcal{IW}_k(n + d_0 + 1, \mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0))$ with $\bar{\mathbf{y}}_\tau(\mathbf{m}_0) = (n\bar{\mathbf{x}}_\tau + r_0\mathbf{m}_0)/(n + r_0)$ and

$$\begin{aligned}
&\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0) = \mathbf{S}_0 + (n-1)\mathbf{S}_\tau + r_0(\boldsymbol{\mu} - \mathbf{m}_0)(\boldsymbol{\mu} - \mathbf{m}_0)^\top + n(\bar{\mathbf{x}}_\tau - \boldsymbol{\mu})(\bar{\mathbf{x}}_\tau - \boldsymbol{\mu})^\top \\
= \;&\mathbf{S}_0 + (n-1)\mathbf{S}_\tau + nr_0 \frac{(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top}{n + r_0} + (n + r_0)(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau((\mathbf{m}_0)))(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top .
\end{aligned}$$

Let $\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0) = \mathbf{S}_0 + (n-1)\mathbf{S}_\tau + nr_0(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))(\mathbf{m}_0 - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top/(n + r_0)$. The application of Sylvester's determinant theorem leads to

$$|\mathbf{V}_\tau(\boldsymbol{\mu}; \mathbf{m}_0, \mathbf{S}_0)| = |\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)|(1 + (n + r_0)(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top \widetilde{\mathbf{V}}_\tau(\mathbf{m}_0, \mathbf{S}_0)^{-1}(\boldsymbol{\mu} - \bar{\mathbf{y}}_\tau(\mathbf{m}_0)))$$

and, hence,

$$
\begin{aligned}
g(\mathbf{m}_0,\mathbf{S}_0) &\propto |\mathbf{S}_0|^{(d_0-k-1)/2}\int_{\boldsymbol{\mu}}|\mathbf{V}_\tau(\boldsymbol{\mu};\mathbf{m}_0,\mathbf{S}_0)|^{-(n+d_0-k)/2}\mathrm{d}\boldsymbol{\mu}\\
&\propto |\mathbf{S}_0|^{(d_0-k-1)/2}|\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0,\mathbf{S}_0)|^{-(n+d_0-k)/2}\\
&\times \int_{\boldsymbol{\mu}}(1+(n+r_0)(\boldsymbol{\mu}-\bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0,\mathbf{S}_0)^{-1}(\boldsymbol{\mu}-\bar{\mathbf{y}}_\tau(\mathbf{m}_0)))^{-(n+d_0-k)/2}\mathrm{d}\boldsymbol{\mu}\\
&\propto |\mathbf{S}_0|^{(d_0-k-1)/2}|\widetilde{\mathbf{V}}_\tau(\mathbf{m}_0,\mathbf{S}_0)|^{-(n+d_0-k-1)/2}\\
&= |\mathbf{S}_0|^{(d_0-k-1)/2}|\mathbf{S}_0+(n-1)\mathbf{S}_\tau|^{-(n+d_0-k-1)/2}\\
&\times \Big(1+nr_0(\mathbf{m}_0-\bar{\mathbf{y}}_\tau(\mathbf{m}_0))^\top(\mathbf{S}_0+(n-1)\mathbf{S}_\tau)^{-1}(\mathbf{m}_0-\bar{\mathbf{y}}_\tau(\mathbf{m}_0))/(n+r_0)\Big)^{-(n+d_0-k-1)/2},
\end{aligned}
$$

where we use Sylvester's determinant theorem for the second time. From the last line, we conclude that $g(\mathbf{m}_0,\mathbf{S}_0)$ is maximized with respect to $\mathbf{m}_0$ at $\hat{\mathbf{m}}_0$ satisfying $\mathbf{m}_0=\bar{\mathbf{y}}_\tau(\mathbf{m}_0)$ independently of $\mathbf{S}_0$ leading to $\hat{\mathbf{m}}_0=\bar{\mathbf{x}}_\tau$.

Taking the logarithms of $g(\mathbf{m}_0,\mathbf{S}_0)$, calculating the matrix derivative with respect to $\mathbf{S}_0$ which is then set to the zero matrix, and substituting $\mathbf{m}_0$ by $\hat{\mathbf{m}}_0$, we get the following matrix equation

$$\frac{d_0-k-1}{2}\mathbf{S}_0^{-1}-\frac{n+d_0-k-1}{2}(\mathbf{S}_0+(n-1)\mathbf{S}_\tau)^{-1}=\mathbf{O}$$

with the solution given by

$$\hat{\mathbf{S}}_0=\frac{(d_0-k-1)(n-1)}{n}\mathbf{S}_\tau.$$

Chapter 3

Bayesian Inference for the Tangent Portfolio

The seminal paper of Markowitz (1952) suggests a simple and intuitive approach for determining the optimal portfolios of risky assets. It allows us to determine the optimal portfolio weights which lead to the lowest risk for a given expected portfolio return. If the asset returns are assumed to follow normal distribution, then this task is equivalent to minimizing the expected quadratic utility of the future wealth. Depending on the level of the risk aversion or on the expected targeted portfolio return, all the resulting portfolios will lie on a hyperbolic efficient frontier in the μ-σ-space. Taking the risk-free asset into account changes the paradigm of the classical Markowitz approach. In this case the efficient portfolios lie on straight line which crosses the vertical axis at the level of the risk-free rate and is tangent to the mean-variance efficient frontier of Markowitz. The line is usually referred to as the capital market line and the tangent point is the tangency portfolio. Every investor holds a portfolio which consists of the tangency portfolio and the risk-free asset, while the proportions are determined by the risk aversion.

In practice, however, the tagency and other portfolios frequently lead to investment strategies with modest profits and high risk. Several approaches were developed to improve the performance. The first strand of research analyses the estimation risk in portfolio weights, which arises if we replace the unknown parameters of the distribution of asset returns with their sample counterparts. If the estimation risk is properly quantified it can be taken into account when constructing estimation-risk-adjusted portfolios. Alternatively one can shrink the optimal portfolio weights to constant target weights. Typically one takes equally weighted portfolio for this purpose. The objective is to minimize an appropriate objective function, usually the utility function of the investor.

The second strand of research uses the Bayesian framework. The Bayesian setting resembles the human way of information utilization. The investors use the past experiences or additional information for decisions at a given time point. These subjective beliefs are reflected in a Bayesian setup using specific prior distributions. The first applications of Bayesian statistics in portfolio analysis were completely based on uninformative or data-based priors, see Winkler (1973), Winkler and Barry (1975). Bawa et al. (1979) provided an excellent review on early examples of Bayesian studies on portfolio choice. These contributions stimulated a steady growth of interest in Bayesian tools for asset allocation problems. Jorion (1986), Kandel and Stambaugh (1996), Barberis (2000), Pástor (2000) used the Bayesian framework to analyze the impact of the underlying asset pricing or predictive model for asset returns on the optimal portfolio choice. Wang (2005), Kan and Zhou (2007), Golosnoy and Okhrin (2007), Golosnoy and Okhrin (2008), Bodnar et al. (2017c)concentrated on shrinkage estimation, which allows to shift the portfolio weights to prespecified values, which reflect the prior beliefs of investors. Brandt (2010) gives a state of the art review of the modern portfolio selection techniques, paying a particular attention to Bayesian approaches.

In this chapter, we consider diffuse and conjugate priors for the parameters of asset returns. In both cases we derive stochastic representation of the posterior distribution of the tangency portfolio and the corresponding first two moments. These results simplify numerical computation of the optimal portfolios and their analysis, since random sampling is required only for simple and standard distribution such as t, N and F. Additionally we provide the asymptotic distribution, which is Gaussian with a simple expression for the covariance matrix. The established results are evaluated within a simulation study, which assesses the coverage probabilities of credible intervals.

The rest of the chapter is structured as follows. Bayesian estimation of the tangency portfolio and main theoretical results are summarized in section 3.1. The results of numerical study are given in section 3.2, while section 3.3 summarizes the chapter. Proofs and additional technical results are contained in section 3.4.

3.1 Bayesian estimation of the tangent portfolio

We consider a portfolio consisting of k assets. The k-dimensional vector of the asset (logarithmic) returns taken at time point t is denoted by $\mathbf{x}_t$. Let w_i be the i-th weight in the portfolio and let $\mathbf{w} = (w_1, \ldots, w_k)^\top$ be the vector of weights. Throughout the paper it is assumed that data drawn from the random vector of asset returns consist of conditionally independent observations which are conditionally normally distributed. That is we assume that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with $\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is a mean vector, $\boldsymbol{\Sigma}$ is a positive definite covariance matrix, and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It is remarkable that only the conditional distribution of the asset returns is assumed to be normal, while the unconditional distribution depends on the priors assigned to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and it is usually a heavy-tailed distribution. Moreover, the observation vectors are unconditionally dependent. These two features, namely heavy tails and time dependence, are, usually, observed in the stochastic behavior of the asset returns.

The aim of this section is to provide a Bayesian analysis of the tangent portfolio (TP) which is an optimal portfolio when the investment into a risk-free asset with return r_f is possible. In the mean-variance space this type of optimal portfolios is determined by a tangent line drawn from the portfolio which consists of the risk-free asset only to the set of optimal portfolios, the so-called efficient frontier, constructed in the case without a risk-free asset. The tangency portfolio (TP) weights are calculated by

$$\mathbf{w}_{TP} = \alpha^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}_k), \tag{3.1}$$

where α is the coefficient of risk aversion which describes the investor's attitude towards risk and $\mathbf{1}_k$ stands for the k-dimensional vector of ones. If the sum of the weights in (3.1) is not equal to one, what is usually observed in practice, then the rest of the investor's wealth is invested into the risk-free rate whose weight is $w_0 = 1 - \mathbf{1}_k^\top \mathbf{w}_{TP}$. Otherwise, if it is normalized so that $\mathbf{1}_k^\top \mathbf{w}_{TP} = 1$, then the TP portfolio coincides with the optimal portfolio that maximizes the Sharpe ratio and it is also known as the market portfolio. This portfolio lies on the intersection of the mean-variance efficient frontier and the capital market line constructed with a risk-free asset.

Obviously the TP weights cannot be calculated since both the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ of the asset return distribution are unknown quantities. They have to be replaced by corresponding estimators using the historical data on asset returns $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$ observed at time points $1, ..., n$. Using these data, the sample estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, namely the sample mean vector and the sample covariance matrix, are constructed and they are given by

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top, \tag{3.2}$$

respectively. Replacing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by $\overline{\mathbf{x}}$ and $\mathbf{S}$ in (3.1) we obtain the sample estimator for the TP weights expressed as

$$\hat{\mathbf{w}}_{TP} = \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k). \tag{3.3}$$

In this section we deal with a more general problem. Namely, the aim is to estimate arbitrary linear combinations of the TP weights. Let $\mathbf{L}$ be a $p \times k$ matrix of constants such that $\text{rank}(\mathbf{L}) =$

$p < k$, and define

$$\boldsymbol{\theta} = \mathbf{L}\mathbf{w}_{TP} = \alpha^{-1}\mathbf{L}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}_k) \tag{3.4}$$

The sample estimator of $\boldsymbol{\theta}$ is then given by

$$\hat{\boldsymbol{\theta}} = \mathbf{L}\hat{\mathbf{w}}_{TP} = \alpha^{-1}\mathbf{L}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k). \tag{3.5}$$

The frequentist distribution of the TP weight and of $\hat{\boldsymbol{\theta}}$ as well as test theory on the TP weights were derived by Bodnar and Okhrin (2011) for $n > k$, whereas Bodnar et al. (2016) extended these results to the case $n < k$.

Here, we deal with the problem of estimating the TP portfolio from the viewpoint of Bayesian statistics. The distributional properties of the TP weights and/or their linear combinations will be presented in terms of the posterior distribution. Thus we obtain not only the point estimator of the weights but the whole distribution. Using the posterior distribution the Bayesian estimate of the TP weights are derived as the posterior mean vector along with their uncertainties which are characterized by the posterior covariance matrix.

The starting point of the Bayesian analysis is the Bayes theorem which relates the posterior distribution of the parameter to the prior distribution and the likelihood function. The latter contains the knowledge about the parameter before the sample is taken. Since the distribution of the asset returns does not directly depend on the TP weights, the posterior distribution of $\mathbf{w}_{TP}$ as well as of $\boldsymbol{\theta}$ is derived from the posterior obtained for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ expressed as

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}) \propto L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3.6}$$

with the likelihood function given by

$$\begin{aligned} L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= L(\mathbf{x}_1, \ldots, \mathbf{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right\} \\ &\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{n}{2}(\overline{\mathbf{x}} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}) - \frac{n-1}{2}\mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}]\right\}. \end{aligned} \tag{3.7}$$

There are several approaches how the prior for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ could be chosen with the diffuse prior and the conjugate prior being the most widely used priors. The diffuse prior belongs to non-informative priors, i.e., it does not incorporate any information for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The diffuse prior is also known as Jeffreys' prior and it is given by

$$\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{k+1}{2}}. \tag{3.8}$$

The second considered prior is the conjugate prior which is an informative one with a normal prior for $\boldsymbol{\mu}$ (conditional on $\boldsymbol{\Sigma}$) and an inverse Wishart prior for $\boldsymbol{\Sigma}$. It is expressed as

$$\begin{aligned}\pi_c(\boldsymbol{\mu}|\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{\kappa_c}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_c)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_c)\right\},\\ \pi_c(\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-\nu_c/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{V}_c\boldsymbol{\Sigma}^{-1}]\right\},\end{aligned}$$

where $\boldsymbol{\mu}_c$ is the prior mean, κ_c is the parameter reflecting the prior precision of $\boldsymbol{\mu}_c$, ν_c is a prior precision on $\boldsymbol{\Sigma}$, and $\mathbf{V}_c$ is a known prior matrix of $\boldsymbol{\Sigma}$. The joint prior for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is then given by

$$\pi_c(\boldsymbol{\mu},\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu_c+1)/2}\exp\left\{-\frac{\kappa_c}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_c)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_c)-\frac{1}{2}\mathrm{tr}[\mathbf{V}_c\boldsymbol{\Sigma}^{-1}]\right\}. \tag{3.9}$$

Both the diffuse and the conjugate priors are successfully applied in finance by Barry (1974), Brown (1976), Klein and Bawa (1976), Frost and Savarino (1986), Rachev et al. (2008), Avramov and Zhou (2010), Sekerke (2015), Bodnar et al. (2017b) among others. The diffuse prior mimics the situation, when the investor has no additional information about the model parameters. The conjugate prior, however, reflects the prior beliefs trough the additional information with the expectations $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$.

In Theorem 9, we derive the stochastic representations of the posterior distributions for $\boldsymbol{\theta}$ under the diffuse prior and the conjugate prior. The stochastic representation is a very powerful tool in multivariate statistics. It plays an important role in the theory of elliptically contoured distributions (c.f., Gupta et al. (2013)) and in Bayesian statistics (see, e.g., Bodnar et al. (2017b)) as well as it is widely used in Monte Carlo studies. In particular, the simulation of the values of the weights is considerably simplified if we use the stochastic representation.

Theorem 9. *Let $\mathbf{x}_1,\ldots,\mathbf{x}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}$ be conditionally independently and identically distributed with $\mathbf{x}_t|\boldsymbol{\mu},\boldsymbol{\Sigma}\sim\mathcal{N}_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Let $\mathbf{L}$ be a $p\times k$ matrix of constants of rank $p<k$, and $\mathbf{1}_k$ denotes the vector of ones. We define*

$$\mathbf{a}_1=\overline{\mathbf{x}}-r_f\mathbf{1}_k,\quad \mathbf{a}_2=\boldsymbol{\mu}_c-r_f\mathbf{1}_k,\quad \textit{and}\quad \mathbf{a}_{12}=\frac{1}{n+\kappa_c}\left(n\mathbf{a}_1+\kappa_c\mathbf{a}_2\right).$$

Then the stochastic representation of the posterior distribution for $\boldsymbol{\theta}=\mathbf{L}\mathbf{w}_{TP}$

(a) under the diffuse prior $\pi_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is given by

$$\boldsymbol{\theta}_d \stackrel{d}{=} \frac{\eta_d}{\alpha}\cdot\mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d+\frac{\sqrt{\eta_d}}{\alpha}\left(\breve{\boldsymbol{\mu}}_d^T\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d\cdot\mathbf{L}\mathbf{S}_d^{-1}\mathbf{L}^\top-\mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d\breve{\boldsymbol{\mu}}_d^T\mathbf{S}_d^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{3.10}$$

with

$$\mathbf{S}_d = \mathbf{S}_d(\breve{\boldsymbol{\mu}}_d) = (n-1)\mathbf{S} + n(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)^\top,$$

where $\eta_d \sim \chi^2_n$, $\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right)$, *and* $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$; *moreover,* η_d, $\breve{\boldsymbol{\mu}}_d$, *and* $\mathbf{z}_0$ *are mutually independent.*

(b) *under the conjugate prior* $\pi_c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta}_c \stackrel{d}{=} \frac{\eta_c}{\alpha} \cdot \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c + \frac{\sqrt{\eta_c}}{\alpha}\left(\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c \cdot \mathbf{L}\mathbf{S}_c^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{3.11}$$

with

$$\begin{aligned}\mathbf{S}_c &= \mathbf{S}_c(\breve{\boldsymbol{\mu}}_c) = \tilde{\mathbf{S}} + (n+\kappa_c)\left(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right)\left(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right)^\top,\\ \tilde{\mathbf{S}} &= (n-1)\mathbf{S} + \mathbf{V}_c - (n+\kappa_c)\mathbf{a}_{12}\mathbf{a}_{12}^\top + (n\mathbf{a}_1\mathbf{a}_1^\top + \kappa_c\mathbf{a}_2\mathbf{a}_2^\top),\end{aligned}$$

where $\eta_c \sim \chi^2_{\nu_c+n-k}$, $\breve{\boldsymbol{\mu}}_c|\mathbf{x} \sim t_k\left(\nu_c+n-2k, \mathbf{a}_{12}, \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}\tilde{\mathbf{S}}\right)$, *and* $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$; *moreover,* η_c, $\breve{\boldsymbol{\mu}}_c$, *and* $\mathbf{z}_0$ *are mutually independent.*

The proof of the theorem is given in the appendix. It is noted that the distribution of the weights is given in terms of a χ^2 random variable, a t-distributed random vector and a standard multivariate normal random vector which are independently distributed. Moreover, only the distribution of $\breve{\boldsymbol{\mu}}_d$ (and $\breve{\boldsymbol{\mu}}_c$) in both stochastic representations depends on data.

To enhance computational efficiency in applications, we rewrite $\mathbf{S}_d^{-1}$ using the Sherman-Morrison formula (see, for example, Meyer (2000, p. 125))

$$\mathbf{S}_d^{-1} = \mathbf{S}_d^{-1}(\breve{\boldsymbol{\mu}}_d) = \frac{1}{n-1}\mathbf{S}^{-1} - \frac{n}{(n-1)^2}\frac{\mathbf{S}^{-1}(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)^\top\mathbf{S}^{-1}}{1 + \frac{n}{n-1}(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)^\top\mathbf{S}^{-1}(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)}.$$

Similarly, we obtain that

$$\mathbf{S}_c^{-1} = \mathbf{S}_c^{-1}(\breve{\boldsymbol{\mu}}_c) = \tilde{\mathbf{S}}_c^{-1} - (n+\kappa_c)\frac{\tilde{\mathbf{S}}_c^{-1}(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12})(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12})^\top\tilde{\mathbf{S}}_c^{-1}}{1 + (n+\kappa_c)(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12})^\top\tilde{\mathbf{S}}_c^{-1}(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12})}.$$

The application of these equalities leads to more computationally efficient stochastic representations of $\boldsymbol{\theta}$ which are stated in Corollary 1.

Corollary 1. *Under the assumptions of Theorem 9, the stochastic representation of* $\boldsymbol{\theta}$

(a) *under the diffuse prior* $\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta} \stackrel{d}{=} \frac{\eta_d}{\alpha} \cdot \mathbf{L}\boldsymbol{\zeta}_d + \frac{\sqrt{\eta_d}}{\alpha}\left(\epsilon_d \cdot \mathbf{L}\boldsymbol{\Upsilon}_d\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_d\boldsymbol{\zeta}_d^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{3.12}$$

with

$$\begin{aligned}
\epsilon_d &= \epsilon_d(Q_d,\mathbf{U}) = \frac{1}{n-1}\mathbf{a}_1^\top\mathbf{S}^{-1}\mathbf{a}_1 + \frac{2}{\sqrt{n-1}}\frac{\sqrt{\frac{k}{n(n-k)}Q_d}}{1+\frac{k}{n-k}Q_d}\mathbf{a}_1^\top\mathbf{S}^{-1/2}\mathbf{U}\\
&\quad+\frac{\frac{k}{n(n-k)}Q_d}{1+\frac{k}{n-k}Q_d} - \frac{\frac{k}{n-k}Q_d}{1+\frac{k}{n-k}Q_d}\frac{1}{n-1}\left(\mathbf{a}_1^\top\mathbf{S}^{-1/2}\mathbf{U}\right)^2,\\
\boldsymbol{\zeta}_d &= \boldsymbol{\zeta}_d(Q_d,\mathbf{U}) = \frac{1}{n-1}\mathbf{S}^{-1}\mathbf{a}_1 + \frac{\sqrt{\frac{k}{n(n-k)}Q_d}}{1+\frac{k}{n-k}Q_d}\frac{1}{\sqrt{n-1}}\mathbf{S}^{-1/2}\mathbf{U}\\
&\quad-\frac{\frac{k}{n-k}Q_d}{1+\frac{k}{n-k}Q_d}\frac{1}{n-1}\mathbf{S}^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}^{-1/2}\mathbf{a}_1,\\
\boldsymbol{\Upsilon}_d &= \boldsymbol{\Upsilon}_d(Q_d,\mathbf{U}) = \frac{1}{n-1}\mathbf{S}^{-1} - \frac{\frac{k}{n-k}Q_d}{1+\frac{k}{n-k}Q_d}\frac{1}{n-1}\mathbf{S}^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}^{-1/2},
\end{aligned}$$

where $\eta_d \sim \chi^2_n$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $Q_d \sim \mathcal{F}(k, n-k)$, *and* $\mathbf{U}$ *is uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η_d, $\mathbf{z}_0$, Q_d, *and* $\mathbf{U}$ *are mutually independent.*

(b) *under the conjugate prior* $\pi_c(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *is given by*

$$\boldsymbol{\theta} \overset{d}{=} \frac{\eta_c}{\alpha}\cdot\mathbf{L}\boldsymbol{\zeta}_c + \frac{\sqrt{\eta_c}}{\alpha}\left(\epsilon_c\cdot\mathbf{L}\boldsymbol{\Upsilon}_c\mathbf{L}^\top - \mathbf{L}\boldsymbol{\zeta}_c\boldsymbol{\zeta}_c^\top\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0, \tag{3.13}$$

with

$$\begin{aligned}
\epsilon_c &= \epsilon_c(Q_c,\mathbf{U}) = \mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12} + 2\frac{\sqrt{\frac{k}{(\kappa_c+n)(\nu_c+n-2k)}Q_c}}{1+\frac{k}{\nu_c+n-2k}Q_c}\mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\\
&\quad+\frac{\frac{k}{(\kappa_c+n)(\nu_c+n-2k)}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c} - \frac{\frac{k}{\nu_c+n-2k}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c}\left(\mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\right)^2,\\
\boldsymbol{\zeta}_c &= \boldsymbol{\zeta}_c(Q_c,\mathbf{U}) = \tilde{\mathbf{S}}^{-1}\mathbf{a}_{12} + \frac{\sqrt{\frac{k}{(\kappa_c+n)(\nu_c+n-2k)}Q_c}}{1+\frac{k}{\nu_c+n-2k}Q_c}\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\\
&\quad-\frac{\frac{k}{\nu_c+n-2k}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c}\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\mathbf{U}^\top\tilde{\mathbf{S}}^{-1/2}\mathbf{a}_{12},\\
\boldsymbol{\Upsilon}_c &= \boldsymbol{\Upsilon}_c(Q_c,\mathbf{U}) = \tilde{\mathbf{S}}^{-1} - \frac{\frac{k}{\nu_c+n-2k}Q_c}{1+\frac{k}{\nu_c+n-2k}Q_c}\tilde{\mathbf{S}}^{-1/2}\mathbf{U}\mathbf{U}^\top\tilde{\mathbf{S}}^{-1/2},
\end{aligned}$$

where $\eta_c \sim \chi^2_{\nu_c+n-k}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $Q_c \sim \mathcal{F}(k, \nu_c+n-2k)$, *and* $\mathbf{U}$ *is uniformly distributed* *on the unit sphere in* $\mathbb{R}^k$*; moreover,* η_c*,* $\mathbf{z}_0$*,* Q_c*, and* $\mathbf{U}$ *are mutually independent.*

In contrast to the stochastic representations given in Theorem 9, the random variables in (3.12) and (3.13) do not depend on data. Consequently, the inverses of the matrices $\mathbf{S}$ and $\tilde{\mathbf{S}}$ have to be calculated only once within the simulation study for a given draw. This would surely speed up the generation of realizations of $\boldsymbol{\theta}$. To this end, it has to be noted that the uniform distribution on a unit sphere in $\mathbb{R}^k$ is not a standard distribution in many statistical packages. However, realizations of $\mathbf{U}$ can easily be obtained from the k-dimensional standard normal vector $\mathbf{Z}$ by using $\mathbf{U} = \mathbf{Z}/\sqrt{\mathbf{Z}^\top \mathbf{Z}}$.

In Theorem 2 we derive the analytical expressions of the Bayesian estimates for $\boldsymbol{\theta}$ calculated under the diffuse prior and the conjugate prior. These expressions are derived as posterior means of $\boldsymbol{\theta}$ which can be calculated by using the results of Corollary 1.

Theorem 10. *Under the assumptions of Theorem 9 the Bayesian estimate for* $\mathbf{w}_{TP}$

(a) under the diffuse prior (3.8) *is given by*

$$\hat{\mathbf{w}}_{TP;d} = E(\mathbf{w}_{TP}|\mathbf{x}) \quad = \quad \frac{1}{\alpha}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f \mathbf{1}_k);$$

(b) under the conjugate prior (3.9) *is given by*

$$\hat{\mathbf{w}}_{TP;c} = E(\mathbf{w}_{TP}|\mathbf{x}) = \frac{\nu_c + n - k - 1}{\alpha}\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12},$$

where $\tilde{\mathbf{S}}$ *and* $\mathbf{a}_{12}$ *are given in Theorem 9.*

The proof is given in section 3.4. From Theorem 10 we observe that the point estimator based on the diffuse prior coincides with the classical estimator. This is consistent with our expectations, since the diffuse prior adds no information, but merely reflect the uncertainty. The conjugate prior is an informative prior. This leads to a new point estimator that is obtained in Theorem 10, which reflects the additional information. The structure of the estimator is of shrinkage-type. The mean vector of excess returns is replaced by the weighted sum of the sample mean return and the prior mean. The weights reflect the precision of both sources of information. The covariance matrix is similarly a weighted sum of the sample and prior information. Thus we shrink the sample parameters towards the priors.

The formulas for the covariance matrix of $\mathbf{w}_{TP}$ under both priors are summarized in Theorem 3 whose proof is given in the appendix.

Theorem 11. *Under the assumptions of Theorem 9 the covariance matrix for* $\mathbf{w}_{TP}$

(a) *under the diffuse prior* (3.8) *is given by*

$$
\begin{aligned}
Var(\mathbf{w}_{TP}|\mathbf{x}) &= \frac{1}{n-1}\alpha^{-2}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)(\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top\mathbf{S}^{-1} \\
+ \quad \alpha^{-2} &\left[\frac{n^2+k-2}{(n-1)n(n+2)} + \frac{n}{(n-1)^2}\frac{k-1}{k}b_d\right]\mathbf{S}^{-1}
\end{aligned}
$$

with $b_d = (\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)$*;*

(b) *under the conjugate prior* (3.9) *is given by*

$$
\begin{aligned}
Var(\mathbf{w}_{TP}|\mathbf{x}) &= (\nu_c + n - k - 1)\alpha^{-2}\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}\mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1} \\
+ \quad \alpha^{-2} &\left[\frac{(\nu_c+n-k)^2+k-2}{(n+\kappa_c)(\nu_c+n-k+2)} + (\nu_c+n-k)\frac{k-1}{k}b_c\right]\tilde{\mathbf{S}}^{-1}
\end{aligned}
$$

with $b_c = \mathbf{a}_{12}^\top\tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}$ *where* $\tilde{\mathbf{S}}$ *and* $\mathbf{a}_{12}$ *are given in Theorem 9.*

Finally, in Theorem 12 we proof that both posterior distributions converge to the same normal distribution as the sample size increases. This results is not surprising and is in line with Bernstein-von-Mises theorem.

Theorem 12. *Under the assumptions of Theorem 9 it holds that*

$$
\sqrt{n}(\mathbf{w}_{TP} - \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k))|\mathbf{x} \xrightarrow{d} \mathcal{N}_k(\mathbf{0}, \mathbf{F}) \tag{3.14}
$$

as $n \to \infty$ *under both the diffuse prior* (3.8) *and the conjugate prior* (3.9) *where*

$$
\mathbf{F} = \alpha^{-2}\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1} + \alpha^{-2}\left[1 + \frac{k-1}{k}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)\right]\breve{\mathbf{S}}^{-1}
$$

where

$$
\breve{\mathbf{x}} = \lim_{n\longrightarrow\infty}\overline{\mathbf{x}} \quad \text{and} \quad \breve{\mathbf{S}} = \lim_{n\longrightarrow\infty}\mathbf{S}.
$$

The proof of the theorem is given in section 3.4. In practice, the asymptotic covariance matrix of $\boldsymbol{w}_{TP}$ is computed by using $\overline{\mathbf{x}}$ and $\mathbf{S}$ instead of $\breve{\mathbf{x}}$ and $\breve{\mathbf{S}}$.

3.2 Simulation Study

In this section we assess the performance of the suggested within a simulation study. We compute the coverage probabilities of credible intervals for the portfolio weights based on the diffuse and conjugate priors suggested in the previous section and compare it to the coverage probability stemming from the asymptotic distribution. Since the posterior distribution cannot be determined explicitly, the quantiles are computed via simulations using the respective stochastic representation. The number of repetitions is set to 10000. To speed up the computations we use the representation in Corollary 1.

The setup of the simulation study is as follows. Without loss of generality we restrict the discussion to the first portfolio weight, i.e. $p = 1$, $\mathbf{L} = \mathbf{e}_1^T$. The riskless rate of return is 0.001. The true expected returns $\boldsymbol{\mu}$ are taken as a unform grid of length k between $-0.01 + r_f$ and $0.01 + r_f$. For the covariance matrix we opt for the AR(1)-type structure $\boldsymbol{\Sigma} = (\rho^{|i-j|})_{i,j=1,..,k}$, where ρ takes values between -1 and 1. Since the dimension of the portfolio is of particular interest we consider $k \in \{5, 10, 20, 30\}$. The sample size n is set to 60, which is a typical value in financial literature and corresponds to roughly two months of daily data or a year of weekly data, respectively. In all considered cases we take the following parameters for the conjugate prior $\nu_c = \kappa_c = n/2$. $\boldsymbol{\mu}_c$ is set equal a uniform grid between 0 and 0.003. $\mathbf{S}_c$ is an identity matrix of a corresponding size. Specifically, the boundaries of the credible intervals are computed using the following procedure:

1. Generate independently
   - Diffuse: $\eta_d \sim \chi^2_n$, conjugate: $\eta_c \sim \chi^2_{\nu_c+n-k}$
   - $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$
   - Diffuse: $Q_d \sim \mathcal{F}(k, n-k)$, conjugate: $Q_c \sim \mathcal{F}(k, \nu_c + n - 2k)$
   - $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k) \to \mathbf{u} = \mathbf{Z}/\sqrt{\mathbf{Z}'\mathbf{Z}}$
2. Compute the vector of weights using (3.12) for the diffuse prior and (3.13) for the conjugate prior and using true parameters.
3. Repeat steps (1) and (2) $B = 10000$ times.
4. Compute the credible intervals bounded by the sample quantiles.

To determine the coverage probabilities we sample the asset returns from the original distribution, estimate the portfolio weights and count the fraction of times the weights are covered by the credible intervals. The results for different dimensions and values of ρ are illustrated in Figure 3.1. We conclude that the diffuse prior leads to the coverage probabilities almost identical to the ones based on the asymptotic distribution and does not depend on the strength of correlation between the assets. The conjugate prior, however, show much higher coverage probabilities especially at the boundaries of ρ. This is reasonable, since higher ρ values induce covariance matrices which are close to singularity. This leads to wider intervals and higher coverage probabilities. Other forms of the correlation structure or other parameters of the conjugate prior might obviously deteriorate these conclusions.

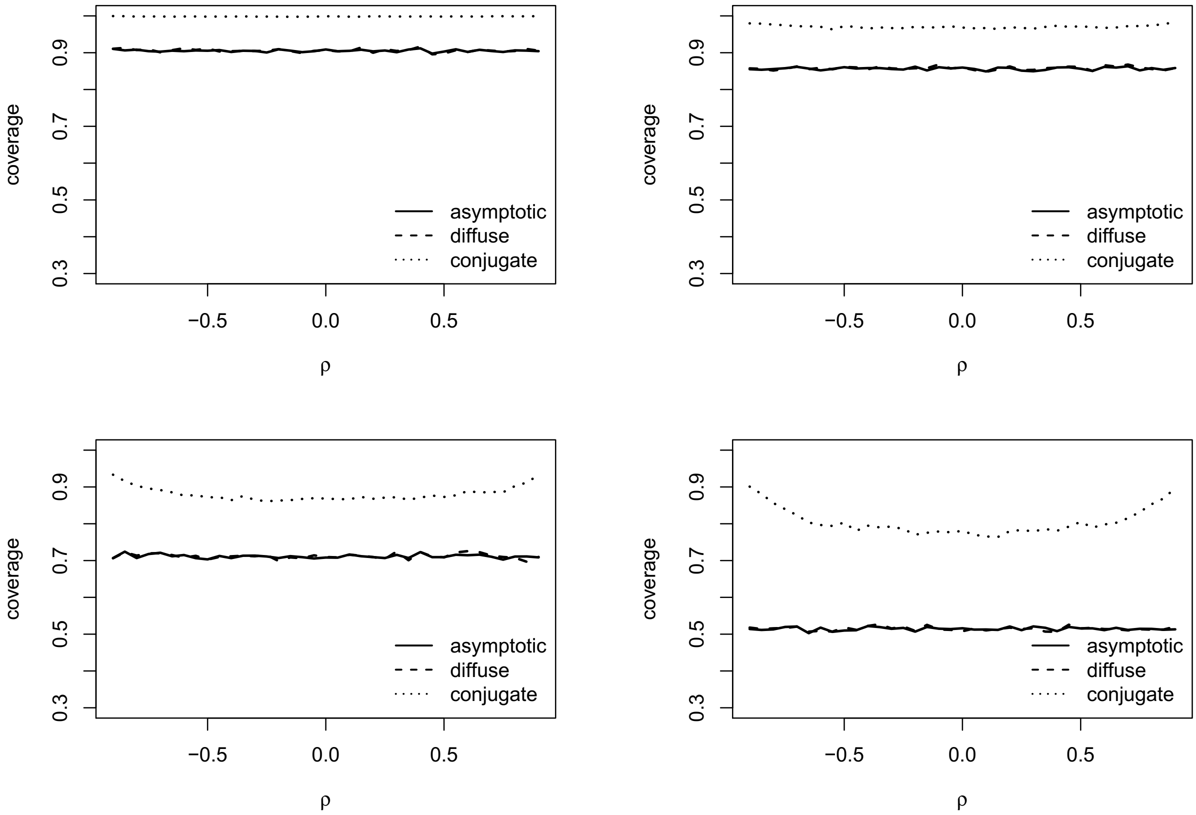

Figure 3.1: Coverage probabilities for k =5, 10 (top) and k =20, 30 (bottom) and 95% level of significance.

3.3 Summary

In this chapter we analyze the tangent portfolio within a Bayesian framework. The suggested approach allows us to incorporate uncertainty about the model parameters quantified as prior beliefs of the investors and to incorporate these into the portfolio decisions. Assuming different priors for the asset returns, we derive the stochastic representation of the posterior distributions of linear combinations of tangent portfolio weights. In particular, we consider non-informative diffuse and informative conjugate priors. Additionally we derive the mean and the variance of the posterior distribution. The results are evaluated within a numerical study, where we assess the coverage probabilities of credible intervals.

3.4 Proofs and Supplementary Material

First, we present an important lemma that is used in the proof of Theorem 9.

Lemma 4. *Assume*

$$\begin{aligned}\boldsymbol{\Xi}|\boldsymbol{\nu},\mathbf{x} &\sim \mathcal{IW}_k(\tau_0,\mathbf{V}_0),\\ \boldsymbol{\nu}|\mathbf{x} &\sim f(\cdot|\mathbf{x}),\end{aligned}$$

where $\mathbf{V}_0=\mathbf{V}_0(\boldsymbol{\nu})$ *and the symbol* $f(\cdot|\mathbf{x})$ *stands for the posterior distribution of* $\boldsymbol{\nu}$*. Let* $\mathbf{M}$ *be a* $p\times k$ *matrix of constants such that rank*$(\mathbf{M})=p\le k$*. Then the stochastic representation of* $\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}$ *is given by*

$$\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}\stackrel{d}{=}\eta\cdot\mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu}+\sqrt{\eta}\left(\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\boldsymbol{\nu}\cdot\mathbf{M}\mathbf{V}_0^{-1}\mathbf{M}^\top-\mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta\sim\chi^2_{\tau_0-k-1}$*,* $\mathbf{z}_0\sim\mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$*, and* $\boldsymbol{\nu}|\mathbf{x}\sim f(\cdot|\mathbf{x})$*; moreover,* $\eta,\mathbf{z}_0$ *and* $\boldsymbol{\nu}$ *are mutually independent.*

Proof. From Theorem 3.4.1 of Gupta and Nagar (2000) we obtain that

$$\boldsymbol{\Xi}^{-1}|\boldsymbol{\nu},\mathbf{x}\sim\mathcal{W}_k(\tau_0-k-1,\mathbf{V}_0^{-1}).$$

Next, we fix $\boldsymbol{\nu}=\boldsymbol{\nu}^*$ and define $\tilde{\mathbf{M}}=(\mathbf{M}^\top,\boldsymbol{\nu}^*)^\top$, $\hat{\mathbf{H}}=\tilde{\mathbf{M}}\boldsymbol{\Xi}^{-1}\tilde{\mathbf{M}}^\top=\left\{\hat{\mathbf{H}}_{ij}\right\}_{i,j=1,2}$ with $\hat{\mathbf{H}}_{11}=\mathbf{M}\boldsymbol{\Xi}^{-1}\mathbf{M}^\top$, $\hat{\mathbf{H}}_{12}=\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}^*$, $\hat{\mathbf{H}}_{21}=\boldsymbol{\nu}^{*T}\boldsymbol{\Xi}^{-1}\mathbf{M}^\top$, and $\hat{H}_{22}=\boldsymbol{\nu}^{*T}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}^*$ as well as $\mathbf{H}=\tilde{\mathbf{M}}\tilde{\mathbf{V}}_0^{-1}\tilde{\mathbf{M}}^\top=\{\mathbf{H}_{ij}\}_{i,j=1,2}$ with $\tilde{\mathbf{V}}_0=\mathbf{V}_0(\boldsymbol{\nu}^*)$, $\mathbf{H}_{11}=\mathbf{M}\tilde{\mathbf{V}}^{-1}\mathbf{M}^\top$, $\mathbf{H}_{12}=\mathbf{M}\tilde{\mathbf{V}}^{-1}\boldsymbol{\nu}^*$, $\mathbf{H}_{21}=\boldsymbol{\nu}^{*T}\tilde{\mathbf{V}}^{-1}\mathbf{M}^\top$, and $H_{22}=\boldsymbol{\nu}^{*T}\tilde{\mathbf{V}}^{-1}\boldsymbol{\nu}^*$.

Since

$$\boldsymbol{\Xi}^{-1}|\boldsymbol{\nu}=\boldsymbol{\nu}^*,\mathbf{x}\sim\mathcal{W}_k(\tau_0-k-1,\tilde{\mathbf{V}}^{-1})$$

and rank$(\tilde{\mathbf{M}})=p+1\le k$, we get from Theorem 3.2.5 in Muirhead (1982) that

$$\hat{\mathbf{H}}|\boldsymbol{\nu}=\boldsymbol{\nu}^*,\mathbf{x}\sim\mathcal{W}_{p+1}(\tau_0-k-1,\mathbf{H}).$$

Moreover, from Theorem 3.2.10 of Muirhead (1982) we obtain that

$$\hat{\mathbf{H}}_{12}|\hat{H}_{22},\boldsymbol{\nu}=\boldsymbol{\nu}^*,\mathbf{x}\sim\mathcal{N}_p(\mathbf{H}_{12}H_{22}^{-1}\hat{H}_{22},\mathbf{H}_{11\cdot 2}\hat{H}_{22}),$$

where $\mathbf{H}_{11\cdot 2}=\mathbf{H}_{11}-\mathbf{H}_{12}H_{22}^{-1}\mathbf{H}_{21}$ is the Schur complement.

Let $\eta = \hat{H}_{22}/H_{22}$. Then the application of Theorem 3.2.8 of Muirhead (1982) leads to

$$\eta|\boldsymbol{\nu} = \boldsymbol{\nu}^*, \mathbf{x} \sim \chi^2_{\tau_0-k-1}.$$

Moreover, since the conditional distribution of η given $\boldsymbol{\nu} = \boldsymbol{\nu}^*$ and $\mathbf{x}$ does not depend on $\boldsymbol{\nu}^*$ and $\mathbf{x}$, it is also the unconditional one as well as η and $\boldsymbol{\nu}$ are independent, i.e. $\eta \sim \chi^2_{\tau_0-k-1}$. Thus, the stochastic representation of $\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}$ is given by

$$\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu} \stackrel{d}{=} \eta \cdot \mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu} + \sqrt{\eta}\left(\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\boldsymbol{\nu}\cdot\mathbf{M}\mathbf{V}_0^{-1}\mathbf{M}^\top - \mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta \sim \chi^2_{\tau_0-k-1}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\boldsymbol{\nu}|\mathbf{x} \sim f(\cdot|\mathbf{x})$; moreover, $\eta, \mathbf{z}_0$ and $\boldsymbol{\nu}$ are mutually independent. This completes the proof of the lemma. □

Proof of Theorem 9. a) Using the expression of the likelihood function (3.7) and the diffuse prior $\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as in (3.8), the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$\begin{aligned}\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}) &\propto L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})\\ &\propto |\boldsymbol{\Sigma}|^{-(n+k+1)/2}\exp\left\{-\frac{n}{2}(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) - \frac{n-1}{2}\mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}]\right\}. \quad (3.15)\end{aligned}$$

Integrating out $\boldsymbol{\Sigma}$ we obtain the marginal posterior for $\boldsymbol{\mu}$ expressed as

$$\begin{aligned}\pi_d(\boldsymbol{\mu}|\mathbf{x}) &\propto \int_{\boldsymbol{\Sigma}>0}|\boldsymbol{\Sigma}|^{-(n+k+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[(n(\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top + (n-1)\mathbf{S})\boldsymbol{\Sigma}^{-1}\right]\right\}d\boldsymbol{\Sigma}\\ &\propto |n(\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top + (n-1)\mathbf{S}|^{-\frac{n}{2}},\end{aligned}$$

where the last equality follows by observing that the function under the integral is the density function of the inverse Wishart distribution with $n+k+1$ degrees of freedom and parameter matrix $n(\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top + (n-1)\mathbf{S}$. The application of Silvester's determinant theorem leads to

$$\pi_d(\boldsymbol{\mu}|\mathbf{x}) \propto \left(1 + \frac{n}{n-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})\right)^{-\frac{n}{2}}$$

which shows that $\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n-k, \overline{\mathbf{x}}, \frac{n-1}{n(n-k)}\mathbf{S}\right)$. Using the properties of the multivariate t-distribution we then get with $\breve{\boldsymbol{\mu}}_d = (\boldsymbol{\mu} - r_f\mathbf{1}_k)$ that

$$\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right) \quad \text{with} \quad \mathbf{a}_1 = \overline{\mathbf{x}} - r_f\mathbf{1}_k.$$

Furthermore, from (5.26) we obtained that $\mathbf{\Sigma}|\breve{\boldsymbol{\mu}}_d, \mathbf{x} \sim \mathcal{IW}_k(n+k+1, \mathbf{S}_d)$ with $\mathbf{S}_d = \mathbf{S}_d(\breve{\boldsymbol{\mu}}) = (n-1)\mathbf{S} + n(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_d - \mathbf{a}_1)^\top$.

Finally, the application of Lemma 8 with $\tau_0 = n+k+1$ and $\mathbf{V}_0 = \mathbf{S}_d$ leads to

$$\boldsymbol{\theta} \stackrel{d}{=} \eta_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d + \sqrt{\eta}\left(\breve{\boldsymbol{\mu}}_d^\top \mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d\breve{\boldsymbol{\mu}}_d^\top \mathbf{S}_d^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta_d \sim \chi^2_n$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, and $\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right)$; moreover, $\eta, \mathbf{z}_0$ and $\breve{\boldsymbol{\mu}}_d$ are mutually independent.

b) The joint posterior for $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ under the conjugate prior (3.9) is given by

$$\pi_c(\boldsymbol{\mu}, \mathbf{\Sigma}|\mathbf{x}) \quad \propto \quad |\mathbf{\Sigma}|^{-(\nu_c+n+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\mathbf{S}_c\mathbf{\Sigma}^{-1}]\right\},$$

where $\breve{\boldsymbol{\mu}}_c = \boldsymbol{\mu} - r_f\mathbf{1}_k$, $\mathbf{a}_1$ and $\mathbf{a}_2$ as in the statement of the theorem and

$$\begin{aligned}\mathbf{S}_c &= \mathbf{S}_c(\breve{\boldsymbol{\mu}}_c) = (n-1)\mathbf{S} + \mathbf{V}_c + n(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_1)(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_1)^\top + \kappa_c(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_2)(\breve{\boldsymbol{\mu}}_c - \mathbf{a}_2)^\top,\\ &= \tilde{\mathbf{S}} + (n+\kappa_c)\left[\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right]\left[\breve{\boldsymbol{\mu}}_c - \mathbf{a}_{12}\right]^\top,\end{aligned}$$

with $\mathbf{a}_{12}$ given in the statement of the theorem and

$$\tilde{\mathbf{S}} = (n-1)\mathbf{S} + \mathbf{V}_c - (n+\kappa_c)\mathbf{a}_{12}\mathbf{a}_{12}^\top + (n\mathbf{a}_1\mathbf{a}_1^\top + \kappa_c\mathbf{a}_2\mathbf{a}_2^\top).$$

Following the proof of part a) of the theorem we get

$$\begin{aligned}\mathbf{\Sigma}|\breve{\boldsymbol{\mu}}_c, \mathbf{x} &\sim \mathcal{IW}_k(\nu_c+n+1, \mathbf{S}_c),\\ \breve{\boldsymbol{\mu}}_c &\sim t_k\left(\nu_c+n-2k, \mathbf{a}_{12}, \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}\tilde{\mathbf{S}}\right).\end{aligned}$$

Finally, the application of Lemma 8 with $\tau_0 = \nu_c+n+1$ and $\mathbf{V}_0 = \mathbf{S}_c$ leads to the statement of the theorem. □

In the proof of Corollary 1 we use the following lemma.

Lemma 5. *Assume that the stochastic representation of* $\mathbf{M}\mathbf{\Xi}^{-1}\boldsymbol{\nu}$ *is given by*

$$\mathbf{M}\mathbf{\Xi}^{-1}\boldsymbol{\nu} \stackrel{d}{=} \eta \cdot \mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu} + \sqrt{\eta}\left(\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\boldsymbol{\nu} \cdot \mathbf{M}\mathbf{V}_0^{-1}\mathbf{M}^\top - \mathbf{M}\mathbf{V}_0^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0,$$

with $\mathbf{V}_0 = \mathbf{V}_0(\boldsymbol{\nu}) = \mathbf{S}_0 + n_0(\boldsymbol{\nu} - \mathbf{b}_0)(\boldsymbol{\nu} - \mathbf{b}_0)^\top$ *and* $\mathbf{M}$ *a* $p \times k$ *matrix of constants such that rank*$(\mathbf{M}) = p \le k$ *where* $\eta \sim \chi^2_{\tau_0}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$, *and* $\boldsymbol{\nu}|\mathbf{x} \sim t_k(d_0, \mathbf{b}_0, \lambda_0\mathbf{S}_0)$*; moreover,* $\eta, \mathbf{z}_0$

and $\boldsymbol{\nu}$ *are mutually independent. Then*

$$\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu} \overset{d}{=} \eta \cdot \mathbf{M}\boldsymbol{\zeta} + \sqrt{\eta}\left(\epsilon \cdot \mathbf{M}\boldsymbol{\Upsilon}\mathbf{M}^\top - \mathbf{M}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{M}^\top\right)^{1/2}\mathbf{z}_0$$

with

$$\begin{aligned}
\epsilon &= \epsilon(Q,\mathbf{U}) = \mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0 + 2\frac{\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ}\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U} \\
&+ \frac{\lambda_0 kQ}{1+n_0\lambda_0 kQ} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\left(\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U}\right)^2, \\
\boldsymbol{\zeta} &= \boldsymbol{\zeta}(Q,\mathbf{U}) = \mathbf{S}_0^{-1}\mathbf{b}_0 + \frac{\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}_0^{-1/2}\mathbf{b}_0, \\
\boldsymbol{\Upsilon} &= \boldsymbol{\Upsilon}(Q,\mathbf{U}) = \mathbf{S}_0^{-1} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}_0^{-1/2},
\end{aligned}$$

where $\eta \sim \chi^2_{\tau_0}$, $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0},\mathbf{I}_p)$, $Q \sim \mathcal{F}(k,d_0)$, *and* $\mathbf{U}$ *uniformly distributed on the unit sphere in* $\mathbb{R}^k$*; moreover,* η, $\mathbf{z}_0$, Q, *and* $\mathbf{U}$ *are mutually independent.*

Proof. Using the Sherman-Morrison formula (see p.125 of Meyer (2000)) we obtain

$$\mathbf{V}_0^{-1} = \mathbf{S}_0^{-1} - n_0\frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)} \tag{3.16}$$

Let

$$\mathbf{U} = \frac{\mathbf{S}_0^{-1/2}(\boldsymbol{\nu}-\mathbf{b}_0)}{\sqrt{(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}} \quad \text{and} \quad Q = \lambda_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)/k. \tag{3.17}$$

Using the facts that $\boldsymbol{\nu}|\mathbf{x} \sim t_k(d_0,\mathbf{b}_0,\lambda_0\mathbf{S}_0)$ and that the multivariate t-distribution belongs to the class of the elliptically contoured distributions, we obtain that $\mathbf{U}$ and Q are independent, and $\mathbf{U}$ is uniformly distributed on the unit sphere in $\mathbb{R}^k$ (see Theorem 2.15 of Gupta et al. (2013)). Moreover, from the properties of the multivariate t-distribution (see p. 19 of Kotz and Nadarajah (2004)), we get that $Q \sim \mathcal{F}(k,d_0)$, i.e., Q has an $\mathcal{F}$-distribution with k and d_0 degrees of freedom.

Hence, the application of the (3.16) and (3.17) leads to

$$\begin{aligned}
\mathbf{V}_0^{-1} &= \mathbf{S}_0^{-1} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}_0^{-1/2},\\
\mathbf{V}_0^{-1}\boldsymbol{\nu} &= \mathbf{S}_0^{-1}\boldsymbol{\nu} - n_0\frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0+\mathbf{b}_0)}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}\\
&= \mathbf{S}_0^{-1}\mathbf{b}_0 + \frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)} - n_0\frac{\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}\mathbf{b}_0}{1+n_0(\boldsymbol{\nu}-\mathbf{b}_0)^\top\mathbf{S}_0^{-1}(\boldsymbol{\nu}-\mathbf{b}_0)}\\
&= \mathbf{S}_0^{-1}\mathbf{b}_0 + \frac{\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\mathbf{S}_0^{-1/2}\mathbf{U}\mathbf{U}^\top\mathbf{S}_0^{-1/2}\mathbf{b}_0,
\end{aligned}$$

and

$$\boldsymbol{\nu}^\top\mathbf{V}_0^{-1}\boldsymbol{\nu} = \mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0 + 2\frac{\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U}\sqrt{\lambda_0 kQ}}{1+n_0\lambda_0 kQ} + \frac{\lambda_0 kQ}{1+n_0\lambda_0 kQ} - \frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\left(\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U}\right)^2.$$

Putting all above together we obtain the statement of the lemma. □

Proof of Corollary 1. The statement of the corollary follows directly from Lemma 9 with $\tau_0 = n$, $n_0 = n$, $d_0 = n-k$, $\lambda_0 = \frac{1}{n(n-k)}$, $\mathbf{b}_0 = \mathbf{a}_1$, $\mathbf{S}_0 = (n-1)\mathbf{S}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{L}$ in the case of the diffuse prior and with $\tau_0 = \nu_c + n - k$, $n_0 = n + \kappa_c$, $d_0 = \nu_c + n - 2k$, $\lambda_0 = \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}$, $\mathbf{b}_0 = \mathbf{a}_{12}$, $\mathbf{S}_0 = \tilde{\mathbf{S}}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{L}$ for the conjugate prior. □

Lemma 6. *Under the assumption of Lemma 9 with $n_0\lambda_0 = 1/d_0$ we get that*

$$E(\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = \tau_0\left(1-\frac{1}{k+d_0}\right)\mathbf{M}^\top\mathbf{S}_0^{-1}\mathbf{b}_0.$$

Proof. Since η, $\mathbf{z}_0$, Q, and $\mathbf{U}$ are independent with $E(\mathbf{z}_0) = \mathbf{0}$, $E(\mathbf{U}) = \mathbf{0}$ and $E(\mathbf{U}\mathbf{U}^\top) = \frac{1}{k}\mathbf{I_k}$, we get

$$E(\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = E(\eta\mathbf{M}\boldsymbol{\zeta}|\mathbf{x}) = \tau_0\left(\mathbf{M}\mathbf{S}_0^{-1}\mathbf{b}_0 - E\left(\frac{n_0\lambda_0 kQ}{1+n_0\lambda_0 kQ}\right)\frac{1}{k}\mathbf{M}\mathbf{S}_0^{-1}\mathbf{b}_0\right).$$

Since $Q \sim \mathcal{F}(k, d_0)$, then from the properties of the $\mathcal{F}$-distribution, we obtain that

$$\frac{\frac{k}{d_0}Q}{1+\frac{k}{d_0}Q} \sim Beta\left(\frac{k}{2}, \frac{d_0}{2}\right).$$

Hence, $E\left(\frac{\frac{k}{d_0}Q}{1+\frac{k}{d_0}Q}\right) = \frac{k}{k+d_0}$ which leads to

$$E(\mathbf{M}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = \tau_0\left(1-\frac{1}{k+d_0}\right)\mathbf{M}\mathbf{S}_0^{-1}\mathbf{b}_0.$$

□

Proof of Theorem 10. The application of Lemma 10 with $\tau_0 = n$, $d_0 = n-k$, $\mathbf{b}_0 = (\overline{\mathbf{x}} - r_f\mathbf{1}_k)$, $\mathbf{S}_0 = (n-1)\mathbf{S}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{l}^\top$ for the diffuse prior and with $\tau_0 = \nu_c + n - k$, $d_0 = \nu_c + n - 2k$, $\mathbf{b}_0 = \mathbf{a}_{12}$, $\mathbf{S}_0 = \tilde{\mathbf{S}}$, $\mathbf{M} = \frac{1}{\alpha}\mathbf{l}^\top$ in the case of the conjugate prior for an arbitrary vector $\mathbf{l}$ leads to

$$E(\mathbf{l}^\top\mathbf{w}_{TP}|\mathbf{x}) = \frac{1}{\alpha}\mathbf{l}^\top\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)$$

and

$$E(\mathbf{l}^\top\mathbf{w}_{TP}|\mathbf{x}) = \frac{\nu_c + n - k - 1}{\alpha}\mathbf{l}^\top\tilde{\mathbf{S}}_c^{-1}\mathbf{a}_{12},$$

respectively. Since the vector $\mathbf{l}$ is arbitrary chosen, we get the statement of the theorem. □

In the proof of Theorem 11 we use the following lemma.

Lemma 7. *Under the assumption of Lemma 9 with $n_0\lambda_0 = 1/d_0$ and $\mathbf{M} = \mathbf{m}^\top : 1\times k$, we get that*

$$\begin{aligned}
Var(\mathbf{m}^\top\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) &= \tau_0(1+\tau_0)\Bigg[\left(1-\frac{2}{k+d_0}+\frac{2}{(k+d_0)(k+d_0+2)}\right)c_{12}^2 \\
&+ \left(\frac{d_0}{n_0(k+d_0)(k+d_0+2)}+\frac{1}{(k+d_0)(k+d_0+2)}c_2\right)c_1\Bigg] \\
&+ \tau_0\Bigg[\left(\frac{k-1}{n_0(k+d_0)}+\left(1-\frac{1}{k}-\frac{1}{k+d_0}+\frac{1}{(k+d_0)(k+d_0+2)}\right)c_2\right)c_1 \\
&+ \frac{2}{(k+d_0)(k+d_0+2)}c_{12}^2\Bigg] - \tau_0^2\left(1-\frac{1}{k+d_0}\right)^2 c_{12}^2,
\end{aligned}$$

where $c_1 = \mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}$, $c_2 = \mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0$, and $c_{12} = \mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{b}_0$.

Proof. It holds that

$$Var(\mathbf{m}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x}) = E(\mathbf{m}^\top\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top\boldsymbol{\Xi}^{-1}\mathbf{m}|\mathbf{x}) - E(\mathbf{m}^\top\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x})^2,$$

where $E(\mathbf{m}\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}|\mathbf{x})$ is given in Lemma 10.

The application of Lemma 9 together with $E(\mathbf{z}_0)=\mathbf{0}$, $E(\mathbf{z}_0\mathbf{z}_0^\top)=\mathbf{I}_p$ and the independence of η, $\mathbf{z}_0$, Q, and $\mathbf{U}$ leads to

$$\begin{aligned}
E(\mathbf{m}^\top\boldsymbol{\Xi}^{-1}\boldsymbol{\nu}\boldsymbol{\nu}^\top\boldsymbol{\Xi}^{-1}\mathbf{m}|\mathbf{x}) &= E(\eta^2)E(\mathbf{m}^\top\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{m}|\mathbf{x})+E(\eta)E\left(\epsilon\mathbf{m}^\top\boldsymbol{\Upsilon}\mathbf{m}-\mathbf{m}^\top\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{m}|\mathbf{x}\right)\\
&= \tau_0(1+\tau_0)E(\mathbf{m}^\top\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{m}|\mathbf{x})+\tau_0E\left(\epsilon\mathbf{m}^\top\boldsymbol{\Upsilon}\mathbf{m}|\mathbf{x}\right),
\end{aligned}$$

where we use that $E(\eta)=\tau_0$ and $E(\eta^2)-E(\eta)=\tau_0(1+\tau_0)$.

Using that $E(\mathbf{U}\mathbf{U}^\top)=\frac{1}{k}\mathbf{I}_k$ and all odd mixed moments of the elements of $\mathbf{U}$ are zero, we get

$$\begin{aligned}
E(\mathbf{m}^\top\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{m}|\mathbf{x}) &= (\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{b}_0)^2+\frac{1}{k}E\left(\frac{\lambda_0kQ}{(1+n_0\lambda_0kQ)^2}\right)\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}\\
&- \frac{2}{k}E\left(\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}\right)(\mathbf{m}\mathbf{S}_0^{-1}\mathbf{b}_0)^2\\
&+ E\left(\left(\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}\right)^2\right)E\left((\mathbf{m}\mathbf{S}_0^{-1/2}\mathbf{U})^2(\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U})^2|\mathbf{x}\right)
\end{aligned}$$

and

$$\begin{aligned}
E\left(\epsilon\mathbf{m}^\top\boldsymbol{\Upsilon}\mathbf{m}|\mathbf{x}\right) &= \mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}+E\left(\frac{\lambda_0kQ}{1+n_0\lambda_0kQ}\right)\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}\\
&- \frac{1}{k}E\left(\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}\right)\mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}\\
&- \frac{1}{k}\mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}-\frac{1}{k}E\left(\frac{\lambda_0kQ}{1+n_0\lambda_0kQ}\right)\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m}\\
&+ E\left(\left(\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}\right)^2\right)E\left((\mathbf{m}\mathbf{S}_0^{-1/2}\mathbf{U})^2(\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U})^2|\mathbf{x}\right).
\end{aligned}$$

Since $\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}$ has a beta distribution with $k/2$ and $d_0/2$ degrees of freedom (see the end of the proof of Lemma 10), we obtain

$$\begin{aligned}
E\left(\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}\right) &= \frac{k}{k+d_0},\\
E\left(\frac{n_0\lambda_0kQ}{1+n_0\lambda_0kQ}\right)^2 &= \frac{2kd_0+k^2(k+d_0+2)}{(k+d_0)^2(k+d_0+2)}=\frac{k(k+2)}{(k+d_0)(k+d_0+2)}.
\end{aligned}$$

Furthermore, using $Q \sim \mathcal{F}(k, d_0)$, we get

$$
\begin{aligned}
E\left[\frac{\lambda_0 k Q}{(1+n_0\lambda_0 k Q)^2}\right] &= \frac{1}{n_0}\int_0^\infty \frac{kt/d_0}{(1+kt/d_0)^2}\frac{1}{B\left(\frac{k}{2},\frac{d_0}{2}\right)}\left(\frac{k}{d_0}\right)^{k/2} t^{k/2-1}\left(1+\frac{k}{d_0}t\right)^{-(k+d_0)/2} dt\\
&= \frac{1}{n_0}\frac{1}{B\left(\frac{k}{2},\frac{d_0}{2}\right)}\int_0^\infty \left(\frac{k}{d_0}\right)^{(k+2)/2} t^{(k+2)/2-1}\left(1+\frac{k}{d_0}t\right)^{-(k+d_0+4)/2} dt\\
&= \frac{1}{n_0}\frac{B\left(\frac{k+2}{2},\frac{d_0+2}{2}\right)}{B\left(\frac{k}{2},\frac{d_0}{2}\right)} = \frac{kd_0}{n_0(k+d_0)(k+d_0+2)},
\end{aligned}
$$

where $B(\cdot,\cdot)$ stands for the beta function (see, Mathai and Provost (1992, p. 256)).

Let $Q_N \sim \chi^2_k$ be independent of $\mathbf{U}$. Then $\sqrt{Q_N}\mathbf{U}$ has a multivariate standard normal distribution, i.e.

$$
\begin{pmatrix}\mathbf{m}^\top\mathbf{S}_0^{-1/2}\\ \mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\end{pmatrix}\sqrt{Q_N}\mathbf{U}|\mathbf{x} \sim \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix}\mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{m} & \mathbf{m}^\top\mathbf{S}_0^{-1}\mathbf{b}_0\\ \mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{m} & \mathbf{b}_0^\top\mathbf{S}_0^{-1}\mathbf{b}_0\end{pmatrix}\right) = \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix}c_1 & c_{12}\\ c_{12} & c_2\end{pmatrix}\right),
$$

where c_1, c_2, and c_{12} are defined in the statement of Lemma 11. Hence,

$$
\begin{aligned}
&E\left((\mathbf{m}\mathbf{S}_0^{-1/2}\mathbf{U})^2(\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U})^2|\mathbf{x}\right) = E\left[\left(\mathbf{m}^\top\mathbf{S}_0^{-1/2}\mathbf{U}\right)^2\left(\mathbf{b}_0^\top\mathbf{S}_0^{-1/2}\mathbf{U}\right)^2|\mathbf{x}\right]\frac{E(Q_N^2)}{E(Q_N^2)}\\
= \ &\frac{E\left[\left(\mathbf{m}^\top\mathbf{S}_0^{-1/2}\sqrt{Q_N}\mathbf{U}\right)^2\left(\mathbf{b}_0^\top\sqrt{Q_N}\mathbf{S}_0^{-1/2}\mathbf{U}\right)^2|\mathbf{x}\right]}{E(Q_N^2)} = \frac{c_1c_2+2c_{12}^2}{k(k+2)},
\end{aligned}
$$

where the last equality follows from the Isserlis' theorem (c.f., Isserlis (1918)).

Thus, we get

$$
\begin{aligned}
E(\mathbf{m}^\top\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\mathbf{m}|\mathbf{x}) &= c_{12}^2 + \frac{1}{k}\frac{kd_0}{n_0(k+d_0)(k+d_0+2)}c_1\\
&- \frac{2}{k}\frac{k}{k+d_0}c_{12}^2 + \frac{k(k+2)}{(k+d_0)(k+d_0+2)}\frac{c_1c_2+2c_{12}^2}{k(k+2)}\\
&= \left(1-\frac{2}{k+d_0}+\frac{2}{(k+d_0)(k+d_0+2)}\right)c_{12}^2\\
&+ \left(\frac{d_0}{n_0(k+d_0)(k+d_0+2)}+\frac{1}{(k+d_0)(k+d_0+2)}c_2\right)c_1
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left(\epsilon \mathbf{m}^\top \boldsymbol{\Upsilon} \mathbf{m}|\mathbf{x}\right) = c_1 c_2 + \frac{1}{n_0}\frac{k}{k+d_0} c_1 - \frac{1}{k}\frac{k}{k+d_0} c_1 c_2 \\
- \; & \frac{1}{k} c_1 c_2 - \frac{1}{k}\frac{1}{n_0}\frac{k}{k+d_0} c_1 + \frac{k(k+2)}{(k+d_0)(k+d_0+2)} \frac{c_1 c_2 + 2c_{12}^2}{k(k+2)} \\
= \; & \frac{2}{(k+d_0)(k+d_0+2)} c_{12}^2 \\
+ \; & \left(\frac{k-1}{n_0(k+d_0)} + \left(1 - \frac{1}{k} - \frac{1}{k+d_0} + \frac{1}{(k+d_0)(k+d_0+2)}\right) c_2\right) c_1 .
\end{aligned}
$$

□

Proof of Theorem 11. For the fixed arbitrary chosen vector $\mathbf{l}$ we apply the results of Lemma 11 with $\tau_0 = n$, $n_0 = n$, $d_0 = n-k$, $\mathbf{b}_0 = (\overline{\mathbf{x}} - r_f \mathbf{1}_k)$, $\mathbf{S}_0 = (n-1)\mathbf{S}$, $\mathbf{m} = \frac{1}{\alpha}\mathbf{l}$ for the diffuse prior and with $\tau_0 = \nu_c + n - k$,$n_0 = n + \kappa_c$, $d_0 = \nu_c + n - 2k$, $\mathbf{b}_0 = \mathbf{a}_{12}$, $\mathbf{S}_0 = \tilde{\mathbf{S}}$, $\mathbf{m} = \frac{1}{\alpha}\mathbf{l}$ in the case of the conjugate prior. This leads to

$$
\begin{aligned}
& Var(\mathbf{l}^\top \mathbf{w}_{TP}|\mathbf{x}) = \frac{n(n+1)}{(n-1)^2}\left[\left(1 - \frac{2}{n} + \frac{2}{n(n+2)}\right) c_{12}^2 + \left(\frac{(n-k)(n-1)}{n^2(n+2)} + \frac{1}{n(n+2)} c_2\right) c_1\right] \\
+ \; & \frac{n}{(n-1)^2}\left[\left(\frac{(k-1)(n-1)}{n^2} + \left(1 - \frac{1}{k} - \frac{1}{n} + \frac{1}{n(n+2)}\right) c_2\right) c_1 + \frac{2}{n(n+2)} c_{12}^2\right] - c_{12}^2 \\
= \; & \frac{1}{n-1} c_{12}^2 + \left[\frac{n^2+k-2}{(n-1)n(n+2)} + \frac{n}{(n-1)^2}\frac{k-1}{k} c_2\right] c_1
\end{aligned}
$$

with $c_1 = \mathbf{l}^\top \mathbf{S}^{-1}\mathbf{l}/\alpha^2$, $c_2 = (\overline{\mathbf{x}} - r_f \mathbf{1}_k)^\top \mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f \mathbf{1}_k)$, $c_{12} = \mathbf{l}^\top \mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f \mathbf{1}_k)/\alpha$ and, similarly,

$$
\begin{aligned}
& Var(\mathbf{l}^\top \mathbf{w}_{TP}|\mathbf{x}) = (\nu_c + n - k - 1) c_{12}^2 \\
+ \; & \left[\frac{(\nu_c + n - k)^2 + k - 2}{(n+\kappa_c)(\nu_c + n - k + 2)} + (\nu_c + n - k)\frac{k-1}{k} c_2\right] c_1
\end{aligned}
$$

with $c_1 = \mathbf{l}^\top \tilde{\mathbf{S}}^{-1}\mathbf{l}/\alpha^2$, $c_2 = \mathbf{a}_{12}^\top \tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}$, and $c_{12} = \mathbf{l}^\top \tilde{\mathbf{S}}^{-1}\mathbf{a}_{12}/\alpha$.

Using the structure of both the variances and the fact that $\mathbf{l}$ is arbitrary chosen, we get the statement of the theorem. □

Proof of Theorem 12. The application of Theorem 9 leads to

$$
\boldsymbol{\theta} \overset{d}{=} \frac{\eta_d}{\alpha} \cdot \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d + \frac{\sqrt{\eta_d}}{\alpha}\left(\breve{\boldsymbol{\mu}}_d^T \mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d \cdot \mathbf{L}\mathbf{S}_d^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_d^{-1}\breve{\boldsymbol{\mu}}_d \breve{\boldsymbol{\mu}}_d^T \mathbf{S}_d^{-1}\mathbf{L}^\top\right)^{1/2} \mathbf{z}_0,
$$

under the diffuse prior, where $\eta_d \sim \chi_n^2$, $\breve{\boldsymbol{\mu}}_d|\mathbf{x} \sim t_k\left(n-k, \mathbf{a}_1, \frac{n-1}{n(n-k)}\mathbf{S}\right)$, and $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ which are mutually independent, and

$$\boldsymbol{\theta} \stackrel{d}{=} \frac{\eta_c}{\alpha} \cdot \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c + \frac{\sqrt{\eta_c}}{\alpha}\left(\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c \cdot \mathbf{L}\mathbf{S}_c^{-1}\mathbf{L}^\top - \mathbf{L}\mathbf{S}_c^{-1}\breve{\boldsymbol{\mu}}_c\breve{\boldsymbol{\mu}}_c^T\mathbf{S}_c^{-1}\mathbf{L}^\top\right)^{1/2}\mathbf{z}_0,$$

where $\eta_c \sim \chi^2_{\nu_c+n-k}$, $\breve{\boldsymbol{\mu}}_c|\mathbf{x} \sim t_k\left(\nu_c + n - 2k, \mathbf{a}_{12}, \frac{1}{(n+\kappa_c)(\nu_c+n-2k)}\tilde{\mathbf{S}}\right)$, and $\mathbf{z}_0 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ which are mutually independent.

Consequently, we get that

$$\sqrt{n}\left(\begin{pmatrix} \eta_d/n \\ \mathbf{z}_0/\sqrt{n} \\ \breve{\boldsymbol{\mu}}_d \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{0} \\ \mathbf{a}_1 \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S} \end{pmatrix}\right)$$

and

$$\sqrt{n}\left(\begin{pmatrix} \eta_c/n \\ \mathbf{z}_0/\sqrt{n} \\ \breve{\boldsymbol{\mu}}_c \end{pmatrix} - \begin{pmatrix} 1 \\ \mathbf{0} \\ \mathbf{a}_1 \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S} \end{pmatrix}\right)$$

as $n \longrightarrow \infty$.

The application of the delta method (c.f., (DasGupta, 2008, Theorem 3.7)) proves that

$$\sqrt{n}(\mathbf{w}_{TP} - \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k))|\mathbf{x} \xrightarrow{d.} \mathcal{N}_k(\mathbf{0}, \mathbf{F}_d)$$

and

$$\sqrt{n}(\mathbf{w}_{TP} - \alpha^{-1}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k))|\mathbf{x} \xrightarrow{d.} \mathcal{N}_k(\mathbf{0}, \mathbf{F}_c),$$

as $n \longrightarrow \infty$ under the diffuse prior and the conjugate prior, respectively.

Finally, using the results of Theorem 11 we get

$$\begin{aligned} \mathbf{F}_d &= \lim_{n\longrightarrow\infty} Var(\sqrt{n}\mathbf{w}_{TP}) = \lim_{n\longrightarrow\infty}\left\{n\frac{1}{n-1}\alpha^{-2}\mathbf{S}^{-1}(\overline{\mathbf{x}} - r_f\mathbf{1}_k)(\overline{\mathbf{x}} - r_f\mathbf{1}_k)^\top\mathbf{S}^{-1}\right. \\ &+ \left.\alpha^{-2}n\left[\frac{n^2+k-2}{(n-1)n(n+2)} + \left(\frac{1}{(n-1)(n+2)} + \frac{n}{(n-1)^2}\frac{k-1}{k}\right)b_d\right]\mathbf{S}^{-1}\right\} \\ &= \alpha^{-2}\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1} + \alpha^{-2}\left[1 + \frac{k-1}{k}\breve{b}_d\right]\breve{\mathbf{S}}^{-1} \end{aligned}$$

where $\breve{b}_d = (\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)$ with $\breve{\mathbf{x}}$ and $\breve{\mathbf{S}}$ defined in the statement of the theorem.

Similarly,

$$\mathbf{F}_c = \alpha^{-2}\breve{\mathbf{S}}^{-1}(\breve{\mathbf{x}} - r_f\mathbf{1}_k)(\breve{\mathbf{x}} - r_f\mathbf{1}_k)^\top\breve{\mathbf{S}}^{-1} + \alpha^{-2}\left[1 + \frac{k-1}{k}\breve{b}_d\right]\breve{\mathbf{S}}^{-1} = \mathbf{F}_d,$$

which completes the proof of the theorem.

□

Chapter 4

Bayesian Inference for the Efficient Frontier

In his seminal paper, Markowitz (1952) opted to choose a portfolio for a given level of the average portfolio return with the smallest risk. This well-known approach was further investigated by Merton (1972) who showed that the set of all portfolios with the smallest risk for a given return level lie on a parabola in the mean-variance space. This parabola is the so-called efficient frontier, described by the expected return and the variance of the global minimum variance portfolio as well as by a slope parameter (Bodnar and Schmid (2009)). While the theoretical properties of the efficient frontier are well examined, the parameters of the efficient frontiers are unknown in practice.

Practitioners have to deal with the sample efficient frontier in practice, where the parameters of the frontier are replaced by their estimates. While there exist a vast literature on possible estimates, see e.g. Lai and Xing (2008) or Brandt (2009), especially the distributional properties of estimates gained attention over the last years. Jobson and Korkie (1980) examined the asymptotic behaviour of the parameter estimators of the efficient frontier, while Jobson (1991) provided the exact distributions for two of the three parameters. Okhrin and Schmid (2006) provided the exact distributions of the weights in the case of the Global Minimum Variance portfolio and Bodnar and Schmid (2009) derived the exact distribution of the whole efficient frontier in the case of a finite sample, therefore extending the asymptotic efficient frontier derived by Jobson and Korkie (1980). Bodnar and Schmid (2008b) and Kan and Smith (2008) independently derived the finite-sample distributions of the estimated parameters of the efficient frontier assuming asset returns to be independent and identically multivariate normally distributed.

While these studies contribute to the vast research on the efficient frontier from the frequentist perspective to assess estimation risk, Bayesian statistics grew popular over the last years for several reasons: Bayesian theory is regarded to resembles the way humans utilize information, especially how investors update their beliefs when facing new events. Most importantly, the Bayesian framework allows to incorporate subjective beliefs on the outcome of a future event which would violate frequentist statistics at its core, see e.g. Avramov and Zhou (2010). While practical information utilization by humans and the resulting problems are studied in different fields of research, the Bayesian framework also allows not to incorporate prior information or subjective beliefs by allowing to access all benefits. These non-informative prior distributions were applied to portfolio theory by Winkler (1973) and Winkler and Barry (1975). Since then, an interest in Bayesian approaches to portfolio theory grew, documented e.g. in Brandt (2009), also fueled by the fact that in finite samples the true distribution of a parameter is accessible in a Bayesian setting and thus asymptotic arguments regarding a distribution in a finite sample are not necessary. For example, Wang (2005), Kan and Zhou (2007) and Bodnar et al. (2017c) focused on shrinkage estimation, allowing for a considerable degree of subjectivity in their portfolio models.

In this chapter, we endow the mean vector and the covariance matrix with the diffuse prior and the conjugate prior. Both priors are regarded as well established in the Bayesian literature and also in research applying Bayesian methods in portfolio selection (see, e.g., Zellner (1971), Klein and Bawa (1976), Frost and Savarino (1986), Rachev et al. (2008), Gelman et al. (2014), Sekerke (2015), Bodnar et al. (2017b)). The diffuse prior is regarded as non-informative, hence ignoring prior knowledge. The conjugate prior allows to incorporate subjective parameterizations which should reflect the practitioners beliefs. Using the resulting posterior distributions and properties of the (inverse) Wishart distribution, we derive stochastic representations for the parameters of the efficient frontier. Stochastic representations are regarded to be powerful tools mostly emphasized in computational statistics, e.g. in Givens and Hoeting (2012) and well established for dealing with elliptically contoured distributions, e.g. by Gupta et al. (2013). In Bayesian statistics, stochastic representations showed to be advantageous as well since they allow to access the posterior distribution of a parameter directly without the need for more complex and ressource-consuming methods like Markov-Chain-Monte-Carlo and circumventing the evaluation of complicated integral expressions. The application of stochastic representations in Bayesian portolio selection was recently demonstrated by Bauder et al. (2017a) and Bauder et al. (2017b). In addition to this, we use the stochastic representations to calculate Bayesian estimates for the parameters as well as their asymptotic distributions. Therefore, this chapter extends the literature regarding Bayesian methods on the one hand as well as providing new insights on the distribution of the parameter estimates of the efficient frontier.

The remainder of this chapter is organized as follows. In the next section, we discussion the estimation of the efficient frontier from the viewpoint of the frequentist statistics, while we derive the stochastic representation from the posterior distribution of the expected return and the variance of the global minimum variance portfolio as well as of the slope parameter of the efficient frontier (Theorem 13) in section 4.2. The Bayesian estimates together with their standard uncertainties are obtained in Theorems 14 and 15. Furthermore, the asymptotic distribution for the three parameters are presented in Theorem 16. In section 4.3, we apply the theoretical findings of section 4.2 to real data consisting of returns on the assets included into the S&P 500. Section 4.4 concludes. Section 4.5 contains all the proofs for the theorems and propositions presented in section 4.2.

4.1 Efficient frontier and its frequentist estimate

Let $\mathbf{x}_t = (x_{t,1}, x_{t,2}, ..., x_{t,k})^\top$ denote a random vector of returns on k assets taken at time point t. Throughout the chapter we assume that the asset returns are infinitely exchangeable and multivariate centered spherically symmetric (cf., Bernardo and Smith (2000, Section 4.4))). This assumption implies that neither the unconditional distribution of the asset returns is normal nor that the asset returns are independently distributed. On the other hand, the imposed assumptions, in particular, imply (see, e.g., Bernardo and Smith (2000, Proposition 4.6)) that the asset returns are independently and identically distributed for a given mean vector $\boldsymbol{\mu}$ and for a given covariance matrix $\boldsymbol{\Sigma}$ with the conditional distribution given by $\mathbf{x}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (k-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$). Moreover, the unconditional distribution of the asset returns appears to be heavy-tailed which is usually observed for financial data.

The quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the parameters of the asset returns distribution where $\boldsymbol{\Sigma}$ is assumed to be a $k \times k$ dimensional positive definite matrix. Denoting the vector of portfolio weights by $\boldsymbol{w}$, i.e., the parts of the investor's wealth invested into each of the selected asset, with $\boldsymbol{w}^\top \mathbf{1} = 1$ where $\mathbf{1}$ the k-dimensional vector of ones, the mean-variance optimization problem is expressed as

$$\min \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \qquad \text{subject to} \qquad \boldsymbol{w}^\top \boldsymbol{\mu} = \mu_0 \qquad \text{and} \qquad \boldsymbol{w}^\top \mathbf{1} = 1 \tag{4.1}$$

for a given level of expected return μ_0. Changing μ_0 we obtain different optimal portfolios. All these portfolios constitute the set of optimal portfolios known as the efficient frontier which is an upper part of a parabola in the mean-variance space (cf., Merton (1972)). This set of optimal portfolios is fully determined by three parameters $R_{GMV} = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} / (\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1})$ and $V_{GMV} = 1/(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1})$ which are the expected return and the variance of the global minimum variance, the optimal portfolio with the smallest variance, and they determine the location of the parabola's vertex in the mean-variance space as well as by the slope parameter of the parabola given by $s = \boldsymbol{\mu}^\top \mathbf{Q} \boldsymbol{\mu}$ with $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}^\top \boldsymbol{\Sigma}{-}1/(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1})$ (c.f., Bodnar and Schmid (2009)). The equation of the efficient frontier is given by

$$(R - R_{GMV})^2 = s(V - V_{GMV})\,. \tag{4.2}$$

Since $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown parameters of the asset return distribution, the efficient frontier cannot be constructed in practice by using (4.2). In practice, the application of its estimate, the sample efficient frontier, is suggested which is expressed as

$$(R - \hat{R}_{GMV})^2 = \hat{s}(V - \hat{V}_{GMV})\,, \tag{4.3}$$

where

$$\hat{R}_{GMV} = \frac{\mathbf{1}^\top \mathbf{S}^{-1}\bar{\mathbf{x}}}{\mathbf{1}^\top \mathbf{S}^{-1}\mathbf{1}}, \quad \hat{V}_{GMV} = \frac{1}{\mathbf{1}^\top \mathbf{S}^{-1}\mathbf{1}}, \quad \hat{s} = \bar{\mathbf{x}}^\top \hat{\mathbf{Q}} \bar{\mathbf{x}} \;\text{ with }\; \hat{\mathbf{Q}} = \mathbf{S}^{-1} - \frac{\mathbf{S}^{-1}\mathbf{1}\mathbf{1}^\top \mathbf{S}^{-1}}{\mathbf{1}^\top \mathbf{S}^{-1}\mathbf{1}} \tag{4.4}$$

are the sample estimates of the three parameters of the efficient frontier with

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \;\text{ and }\; \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top\,. \tag{4.5}$$

The notations $\bar{\mathbf{x}}$ and $\mathbf{S}$ denote the sample estimates of the expected return vector and the covariance matrix of the asset returns based on the sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

The distributional properties of the frequentist estimate of the efficient frontier have been discussed in the length of literature studies. The asymptotic behavior of the sample efficient frontier in the case of finite dimension k and normally distributed asset returns was investigated by Jobson and Korkie (1980) and Jobson (1991), whereas Bodnar and Schmid (2009) and Kan and Smith (2008) studied the finite sample distributional properties of $\{\hat{R}_{GMV}, \hat{V}_{GMV}, \hat{s}\}$. Asymptotic results under the double asymptotic regime, i.e. for large k, are available in Bodnar et al. (2017a), while Bodnar and Gupta (2009) presented finite-sample results for elliptically distributed asset returns. Nevertheless, Basak et al. (2005) and Siegel and Woodgate (2007) showed that the sample efficient frontier overestimates the true location of the efficient frontier in the mean-variance space. In order to correct this overoptimism Kan and Smith (2008) derived improved estimates for the parameters of the efficient frontier, while Bodnar and Bodnar (2010) constructed an unbiased estimate of the whole efficient frontier.

4.2 Bayesian inference for the efficient frontier

4.2.1 Statistical model and priors

In the following section, we deal with the problem of estimating the efficient frontier from the viewpoint of Bayesian statistics. Endowing the parameters of the asset returns with a prior distribution, the posterior for the parameters of the efficient frontier will be derived. This finding allows us to characterize the location of the efficient frontier in the mean-variance space as well as to provide point estimates obtained under several loss functions. To this end, we point out that the asset returns are neither assumed to be normally nor independently distributed. The assumption of independence is replaced by the weaker one of exchangeability, while instead of normality it is assumed that the asset returns are centered spherically symmetric.

The Bayes theorem states that

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{x}) \propto L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})\,, \tag{4.6}$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ stands for the $p \times n$ data matrix. The symbol $L(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the likelihood function. Using the expression of the posterior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as in (4.6), we then derive the posterior distribution of the three parameters of the efficient frontier, namely $\pi(R_{GMV}, V_{GMV}, s|\mathbf{x})$. The choice of the prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is an important step in the Bayesion decision process. Several priors for the mean vector and covariance matrix of the asset returns have been suggested in the literature (see, e.g., Barry (1974), Brown (1976), Klein and Bawa (1976), Frost and Savarino (1986), Rachev et al. (2008), Avramov and Zhou (2010), Sekerke (2015)) with the recent paper of Bodnar et al. (2017b) summarizing these results. We make use of the Jeffreys non-informative prior and a conjugate informative prior which are widely applied when Bayesian inferences for optimal portfolios are discussed.

The Jeffreys non-informative prior is also known as the diffuse prior and it is given by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \propto \quad |\boldsymbol{\Sigma}|^{-(k+1)/2} \tag{4.7}$$

while the conjugate prior is expressed as

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \quad \sim \quad \mathcal{N}_k\left(\mathbf{m}_0, \frac{1}{r_0}\boldsymbol{\Sigma}\right) \tag{4.8}$$

$$\boldsymbol{\Sigma} \quad \sim \quad \mathcal{IW}_k(d_0, \mathbf{S}_0). \tag{4.9}$$

Here $\mathbf{m}_0$, r_0, d_0, and $\mathbf{S}_0$ are additional model parameters known as hyperparameters. The symbol $\mathcal{IW}_k(d_0, \mathbf{S}_0)$ stands for the inverse Wishart distribution with d_0 degrees of freedom and parameter matrix $\mathbf{S}_0$. The prior mean $\mathbf{m}_0$ reflects our prior expectation about the expected asset returns, while $\mathbf{S}_0$ presents the prior beliefs about the covariance matrix. The other two hyperparameters r_0 and d_0 are known as precision parameters for $\mathbf{m}_0$ and $\mathbf{S}_0$, respectively. Note that the prior (4.8)-(4.9) corresponds to the well-known conjugate normal-inverse-Wishart model as discussed by, e.g., Gelman et al. (2014). In this case the posterior is accessible in an analytical form and, moreover, it has the same distribution as the prior with updated hyperparameters.

4.2.2 Posterior distribution

In Proposition 3, we present the marginal posterior of $\boldsymbol{\mu}$ as well as the conditional posterior of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$. These results will be later used in the derivation of Bayesian estimates for the three parameters of the efficient frontier. Let $t_k(d, \mathbf{a}, \mathbf{A})$ denote the k-dimensional t-distribution with d degrees of freedom, location vector $\mathbf{a}$ and dispersion matrix $\mathbf{A}$. In the case of $k = 1$, $\mathbf{a} = 0$, and $\mathbf{A} = 1$ we use the simplified notation t_d instead. The following proposition is taken from Bauder et al. (2017a).

Proposition 3. *Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be conditionally independently distributed with $\mathbf{X}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = 1, ..., n$ with $n > k$. Then:*

(a) Under the diffuse prior (4.7), the marginal posterior distribution of $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n-k, \overline{\mathbf{x}}_d, \frac{1}{n(n-k)}\mathbf{S}_d\right) \quad \text{with} \quad \overline{\mathbf{x}}_d = \overline{\mathbf{x}} \quad \text{and} \quad \mathbf{S}_d = (n-1)\mathbf{S}.$$

The conditional posterior distribution of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ is expressed as

$$\boldsymbol{\Sigma}|\boldsymbol{\mu}, \mathbf{x} \sim \mathcal{IW}_k(n+k+1, \mathbf{S}_d^*(\boldsymbol{\mu})) \quad \text{with} \quad \mathbf{S}_d^*(\boldsymbol{\mu}) = \mathbf{S}_d + n(\boldsymbol{\mu} - \overline{\mathbf{x}}_d)(\boldsymbol{\mu} - \overline{\mathbf{x}}_d)^\top.$$

(b) Under the conjugate prior (4.8) *and* (4.9), *the marginal posterior distribution of $\boldsymbol{\mu}$ is given by*

$$\boldsymbol{\mu}|\mathbf{x} \sim t_k\left(n+d_0-2k, \overline{\mathbf{x}}_c, \frac{1}{(n+r_0)(n+d_0-2k)}\mathbf{S}_c\right) \quad \text{with}$$

$$\overline{\mathbf{x}}_c = \frac{n\overline{\mathbf{x}} + r_0\mathbf{m}_0}{n+r_0} \quad \text{and} \quad \mathbf{S}_c = \mathbf{S}_d + \mathbf{S}_0 + nr_0\frac{(\mathbf{m}_0 - \overline{\mathbf{x}}_c)(\mathbf{m}_0 - \overline{\mathbf{x}}_c)^\top}{n+r_0}.$$

The conditional posterior distribution of $\boldsymbol{\Sigma}$ given $\boldsymbol{\mu}$ is expressed as

$$\begin{aligned}\boldsymbol{\Sigma}|\boldsymbol{\mu}, \mathbf{x} &\sim IW_k(n+d_0+1, \mathbf{S}_c^*(\boldsymbol{\mu})) \quad \text{with} \\ \mathbf{S}_c^*(\boldsymbol{\mu}) &= \mathbf{S}_c + (n+r_0)(\boldsymbol{\mu} - \overline{\mathbf{x}}_c)(\boldsymbol{\mu} - \overline{\mathbf{x}}_c)^\top.\end{aligned}$$

In order to assess the risk associated with estimating R_{GMV}, V_{GMV}, and s, we present the joint posterior distribution of these three parameters of the efficient frontier in Theorem 13. This is achieved by deriving their stochastic representations, endowing the parameters with their diffuse and conjugate priors. These findings are later used in the calculation of Bayesian estimates for the three parameters of the efficient frontier (Theorem 14) as well as for their covariance matrix (Theorem 15). It is noted that the application of the stochastic representation to describe the distribution of random quantities has been used both in conventional statistics (see, e.g., Gupta et al. (2013)) and in Bayesian statistics (c.f., Bodnar et al. (2017b)). Later on, the symbol "$\stackrel{d}{=}$" denotes equality in distribution. The proof of Theorem 13 is presented in section 4.5.

Theorem 13. *Under the assumption of Proposition 3 we get:*

(a) Under the diffuse prior (4.7), the stochastic representation is given by

$$
\begin{aligned}
V_{GMV} &\stackrel{d}{=} V_{GMV;d}\left(1+\frac{\psi_1^2}{\phi+\varphi+\psi_2^2}\right)\tau^{-1}, &(4.10)\\
R_{GMV} &\stackrel{d}{=} R_{GMV;d}-\sqrt{V_{GMV;d}}\frac{\sqrt{s_d}\psi_2-\sqrt{\phi/n}}{\phi+\varphi+\psi_2^2}\psi_1\\
&+\sqrt{V_{GMV;d}}\sqrt{1+\frac{\psi_1^2}{\phi+\varphi+\psi_2^2}}\frac{\eta}{\sqrt{\tau}}\sqrt{s_d+\frac{1}{n}-\frac{(\sqrt{s_d}\psi_2-\sqrt{\phi/n})^2}{\phi+\varphi+\psi_2^2}}, &(4.11)\\
s &\stackrel{d}{=} \xi\left(s_d+\frac{1}{n}-\frac{(\sqrt{s_d}\psi_2-\sqrt{\phi/n})^2}{\phi+\varphi+\psi_2^2}\right)\\
&= \xi\left(\frac{\varphi\left(s_d+\frac{1}{n}\right)+\left(\sqrt{s_d}\sqrt{\phi}+\frac{\psi_2}{\sqrt{n}}\right)^2}{\phi+\varphi+\psi_2^2}\right) &(4.12)
\end{aligned}
$$

with

$$
R_{GMV;d}=\frac{\mathbf{1}^\top\mathbf{S}_d^{-1}\overline{\mathbf{x}}_d}{\mathbf{1}^\top\mathbf{S}_d^{-1}\mathbf{1}},\quad V_{GMV;d}=\frac{1}{\mathbf{1}^\top\mathbf{S}_d^{-1}\mathbf{1}},\quad \text{and}\quad s_d=\overline{\mathbf{x}}_d^\top\mathbf{S}_d^{-1}\overline{\mathbf{x}}_d-\frac{(\mathbf{1}^\top\mathbf{S}_d^{-1}\overline{\mathbf{x}}_d)^2}{\mathbf{1}^\top\mathbf{S}_d^{-1}\mathbf{1}}, \tag{4.13}
$$

where $\xi\sim\chi^2_{n-1}$, $\tau\sim\chi^2_n$, $\varphi\sim\chi^2_{k-2}$, $\phi\sim\chi^2_{n-k}$, $\eta\sim\mathcal{N}(0,1)$, and $\boldsymbol{\psi}=(\psi_1,\psi_2)^\top\sim\mathcal{N}_2(\mathbf{0},\mathbf{I})$, and they are mutually independently distributed.

(b) Under the conjugate prior (4.8) and (4.9), the stochastic representation is given by

$$\begin{aligned} V_{GMV} &\overset{d}{=} V_{GMV;c}\left(1+\frac{\psi_1^2}{\phi+\varphi+\psi_2^2}\right)\tau^{-1}, \qquad (4.14)\\ R_{GMV} &\overset{d}{=} R_{GMV;c}-\sqrt{V_{GMV;c}}\frac{\sqrt{s_c}\psi_2-\sqrt{\phi/(n+r_0)}}{\phi+\varphi+\psi_2^2}\psi_1\\ &+ \sqrt{V_{GMV;c}}\sqrt{1+\frac{\psi_1^2}{\phi+\varphi+\psi_2^2}}\frac{\eta}{\sqrt{\tau}}\\ &\cdot \sqrt{s_c+\frac{1}{n+r_0}-\frac{(\sqrt{s_c}\psi_2-\sqrt{\phi/(n+r_0)})^2}{\phi+\varphi+\psi_2^2}}, \qquad (4.15)\\ s &\overset{d}{=} \xi\left(s_c+\frac{1}{n+r_0}-\frac{(\sqrt{s_c}\psi_2-\sqrt{\phi/(n+r_0)})^2}{\phi+\varphi+\psi_2^2}\right)\\ &= \xi\left(\frac{\varphi\left(s_c+\frac{1}{n+r_0}\right)+\left(\sqrt{s_c}\sqrt{\phi}+\frac{\psi_2}{\sqrt{n+r_0}}\right)^2}{\phi+\varphi+\psi_2^2}\right) \qquad (4.16)\end{aligned}$$

with

$$R_{GMV;c}=\frac{\mathbf{1}^\top\mathbf{S}_c^{-1}\overline{\mathbf{x}}_c}{\mathbf{1}^\top\mathbf{S}_c^{-1}\mathbf{1}},\quad V_{GMV;c}=\frac{1}{\mathbf{1}^\top\mathbf{S}_c^{-1}\mathbf{1}},\quad \text{and}\quad s_c=\overline{\mathbf{x}}_c^\top\mathbf{S}_c^{-1}\overline{\mathbf{x}}_c-\frac{(\mathbf{1}^\top\mathbf{S}_c^{-1}\overline{\mathbf{x}}_c)^2}{\mathbf{1}^\top\mathbf{S}_c^{-1}\mathbf{1}}, \qquad (4.17)$$

where $\xi\sim\chi^2_{n+d_0-k-1}$, $\tau\sim\chi^2_{n+d_0-k}$, $\varphi\sim\chi^2_{k-2}$, $\phi\sim\chi^2_{n+d_0-2k}$, $\eta\sim\mathcal{N}(0,1)$, *and* $\boldsymbol{\psi}=(\psi_1,\psi_2)^\top\sim\mathcal{N}_2(\mathbf{0},\mathbf{I})$, *and they are mutually independently distributed.*

Theorem 13 possesses several important applications which are summarized in the following subsections. Furthermore, from the proof of Lemma 11 (see section 4.5) we get that V_{GMV} and s are independently distributed under both the diffuse prior and the conjugate prior since

$$\phi+\varphi+\psi_2^2 \quad\text{and}\quad \frac{\sqrt{s_c}\psi_2-\sqrt{\phi/n}}{\sqrt{\phi+\varphi+\psi_2^2}}$$

are independent with $\varsigma=\phi+\varphi+\psi_2^2$ being χ^2_{n-1} under the diffuse prior and $\chi^2_{n+d_0-k-1}$ under the conjugate prior.

4.2.3 Point estimation

There are several ways how a Bayes estimate of a parameter could be obtained from its posterior. The most commonly procedures are to apply the posterior mean, the posterior median, or the posterior mode. Each of these point estimates corresponds to a different loss function used in its derivation (cf. Bernardo and Smith (2000, Proposition 5.2)). The posterior means under both the diffuse prior and the conjugate prior could be calculated analytically using the stochastic representation of the three parameters of the efficient frontier as given in Theorem 13. We present these results in Theorem 14 with the proof given in section 4.5.

Theorem 14. *Let the assumptions of Proposition 3 hold and assume that $n \geq 3$. Then:*

(a) Under the diffuse prior (4.7), the posterior means are given by

$$\hat{V}_{GMV;d} = \mathbb{E}(V_{GMV}|\mathbf{x}) = \frac{1}{n-3} V_{GMV;d}, \tag{4.18}$$

$$\hat{R}_{GMV;d} = \mathbb{E}(R_{GMV}|\mathbf{x}) = R_{GMV;d}, \tag{4.19}$$

$$\hat{s}_d = \mathbb{E}(s|\mathbf{x}) = (n-2)s_d + \frac{k-1}{n}. \tag{4.20}$$

(b) Under the conjugate prior (4.8) and (4.9), the posterior means are given by

$$\hat{V}_{GMV;c} = \mathbb{E}(V_{GMV}|\mathbf{x}) = \frac{1}{n+d_0-k-3} V_{GMV;c}, \tag{4.21}$$

$$\hat{R}_{GMV;c} = \mathbb{E}(R_{GMV}|\mathbf{x}) = R_{GMV;c}, \tag{4.22}$$

$$\hat{s}_c = \mathbb{E}(s|\mathbf{x}) = (n+d_0-k-2)s_c + \frac{k-1}{n+r_0}. \tag{4.23}$$

The other two Bayesian estimates, the posterior median and the posterior mode, are obtained via simulations by applying the stochastic representations of Theorem 13. Namely, the following algorithm could be used:

(1) Generate independently $\varphi \sim \chi^2_{k-2}$, $\eta \sim \mathcal{N}(0,1)$, and $\boldsymbol{\psi} \sim \mathcal{N}_2(\mathbf{0},\mathbf{I})$, as well as $\phi \sim \chi^2_{n-k}$, $\xi \sim \chi^2_{n-1}$, and $\tau \sim \chi^2_n$ under the diffuse prior (4.7) or $\phi \sim \chi^2_{n+d_0-2k}$, $\xi \sim \chi^2_{n+d_0-k-1}$, and $\tau \sim \chi^2_{n+d_0-k}$ under the conjugate prior (4.8) and (4.9);

(2) Using data matrix $\mathbf{x}$ and prior hyperparameters, compute $R_{GMV;d}$, $V_{GMV;d}$ and s_d under the diffuse prior (4.7) or $R_{GMV;c}$, $V_{GMV;c}$ and s_c under the conjugate prior (4.8) and (4.9);

(3) Apply the stochastic representation of Theorem 13 to get realizations of the three parameter of the efficient frontier;

(4) Repeat steps (1) and (3) B times

As a result, we get a sample from the posterior distribution of the three parameters of the efficient frontier $(R^{(b)}_{GMV}, V^{(b)}_{GMV}, s^{(b)})$, $b = 1, ..., B$ from which the sample median/mode is used as a proxy for the posterior median/mode.

The posterior variances and covariances can also be calculated analytically using the results of Theorem 13 and they are given by

Theorem 15. *Let the assumptions of Proposition 3 hold and assume that $n \geq 5$. Then the posterior covariances between R_{GMV}, V_{GMV}, and s are all zeros under both the diffuse prior (4.7) and the conjugate prior (4.8) and (4.9). We further get*

(a) Under the diffuse prior (4.7), the posterior variances are given by

$$\begin{aligned}
u_d^2(V_{GMV}) &= \mathbb{V}ar(V_{GMV}|\mathbf{x}) = 2V_{GMV;d}^2 \frac{1}{(n-3)^2(n-5)},\\
u_d^2(R_{GMV}) &= \mathbb{V}ar(R_{GMV}|\mathbf{x}) = V_{GMV;d}\left(s_d + \frac{1}{n}\right)\frac{1}{n-3},\\
u_d^2(s) &= \mathbb{V}ar(s|\mathbf{x}) = 2(n-2)\left(s_d^2 + 2\frac{s_d}{n}\right) + \frac{2k-2}{n^2}.
\end{aligned}$$

(b) Under the conjugate prior (4.8) and (4.9), the posterior variances are given by

$$\begin{aligned}
u_c^2(V_{GMV}) &= \mathbb{V}ar(V_{GMV}|\mathbf{x}) = 2V_{GMV;c}^2 \frac{1}{(n+d_0-k-3)^2(n+d_0-k-5)},\\
u_c^2(R_{GMV}) &= \mathbb{V}ar(R_{GMV}|\mathbf{x}) = V_{GMV;c}\left(s_c + \frac{1}{n+r_0}\right)\frac{1}{n+d_0-k-3},\\
u_c^2(s) &= \mathbb{V}ar(s|\mathbf{x}) = 2(n+d_0-k-2)\left(s_c^2 + 2\frac{s_c}{n+r_0}\right) + \frac{2k-2}{(n+r_0)^2}.
\end{aligned}$$

4.2.4 Asymptotic distribution

In Theorem 16 we prove that the posterior distribution of the three parameters of the efficient frontier converges to the same normal distribution under both the diffuse prior and the conjugate prior.

Theorem 16. *Under the assumption of Proposition 3, it holds that*

$$\sqrt{n}\begin{pmatrix} \sqrt{1/2}\,nV_{GMV;d}^{-1}\left(V_{GMV} - \frac{V_{GMV;d}}{n}\right) \\ \sqrt{n}V_{GMV;d}^{-1/2}\left(1+ns_d\right)^{-1/2}\left(R_{GMV} - R_{GMV;d}\right) \\ \left(2n^2s_d^2 + 4ns_d\right)^{-1/2}\left(s - ns_d\right) \end{pmatrix} \xrightarrow{d} \mathcal{N}_3(\mathbf{0}, \mathbf{I}) \quad as \quad n \longrightarrow \infty \tag{4.24}$$

under the diffuse prior (4.7) and

$$\sqrt{n}\begin{pmatrix} \sqrt{1/2}\,nV_{GMV;c}^{-1}\left(V_{GMV} - \frac{V_{GMV;c}}{n}\right) \\ \sqrt{n}V_{GMV;c}^{-1/2}\left(1+ns_c\right)^{-1/2}\left(R_{GMV} - R_{GMV;c}\right) \\ \left(2n^2s_c^2 + 4ns_c\right)^{-1/2}\left(s - ns_c\right) \end{pmatrix} \xrightarrow{d} \mathcal{N}_3(\mathbf{0}, \mathbf{I}) \quad as \quad n \longrightarrow \infty \tag{4.25}$$

under the conjugate prior (4.8) and (4.9).

The proof of Theorem 16 is given in section 4.5. Assume that the following two limits almost surely exist and are equal to

$$\lim_{n\longrightarrow\infty} \overline{\mathbf{x}} = \breve{\mathbf{x}} \quad \text{and} \quad \lim_{n\longrightarrow\infty} \mathbf{S} = \breve{\mathbf{S}}.$$

Then it directly follows almost surely that

$$\lim_{n\longrightarrow\infty} \overline{\mathbf{x}}_d = \lim_{n\longrightarrow\infty} \overline{\mathbf{x}}_c = \breve{\mathbf{x}} \quad \text{and} \quad \lim_{n\longrightarrow\infty} \frac{\mathbf{S}_d}{n} = \lim_{n\longrightarrow\infty} \frac{\mathbf{S}_c}{n+r_0} = \breve{\mathbf{S}}.$$

Consequently, we get almost surely

$$\begin{aligned}
&\lim_{n\longrightarrow\infty} \frac{V_{GMV;d}}{n} = \lim_{n\longrightarrow\infty} \frac{V_{GMV;c}}{n} = \breve{V}_{GMV},\\
&\lim_{n\longrightarrow\infty} R_{GMV;d} = \lim_{n\longrightarrow\infty} R_{GMV;c} = \breve{R}_{GMV},\\
&\lim_{n\longrightarrow\infty} ns_d = \lim_{n\longrightarrow\infty} ns_c = \breve{s}.
\end{aligned}$$

Hence, the results of Theorem 16 are in line with the Bernstein-von Mises theorem (c.f., Bernardo and Smith (2000)) which shows under some regularity conditions that the posterior converges to the normal distribution independently of the prior used when the sample size tends to infinity. The differences which are present on the left hand-sides of (4.24) and (4.25) are nothing else as the finite-sample corrections of the obtained asymptotic distributions.

4.3 Empirical illustration

In this section we present how the methods proposed in the previous section work using real data. The focus lies on the issues most important to the practitioner: the precision of the efficient-frontier estimation, how precise is the amount of risk associated with a specific return and how precise this risk can be determined. Using Theorem 1, these tasks can be tackled easily by assessing the posterior distributions of V_{GMV}, R_{GMV} and s_{GMV} directly. The steps needed are described in the previous section. But in addition to this advantage, the stochastic representations of Theorem 1 are computationally highly efficient since only sampling from well known distributions is necessary and the sample covariance matrix has to be inverted only once. By choosing a sufficiently high number of draws B, a high numerical precision can be achieved.

We use weekly returns from a collection of assets ranging from five to 50 assets of the S&P500, representing a broad range of branches within the United States of America. Weekly returns are a compromise between actuality of the parameters and fulfilling the assumption of conditionally normally distributed returns. Weekly returns do not deviate severely from this assumption but are sufficiently actual. The parameters are estimated with sample sizes of $n \in \{52, 78, 104, 130\}$, corresponding from one year up to two and a half years of weekly data. The data ends at 11.07.2017, covering a period of relative stability for $n = 52$ and $n = 78$ and with two drops in August 2015 and the early weeks of 2016 which are covered by $n = 104$ and $n = 130$. The portfolios are of four different sizes $k \in \{5, 10, 20, 50\}$ in order to assess the behaviour of the parameters' distributions for different sample sizes as well as for different numbers of assets. For this we chose $r_0 = n/10$ and $d_0 = \lfloor n/10 \rfloor$. $\boldsymbol{\mu}_0$ and $\mathbf{S}_0$ are set to the mean vector and covariance matrix using the n data-points prior to this time point, where the regular portfolio estimates are estimated from.

Figure 4.1 and 4.2 show the exact 95% credible sets for V and R constructed by employing the diffuse (black line) and the conjugate (grey line) priors. The asymptotic credible sets finitely-sample corrected as in Theorem 16 are drawn by using the corresponding dash lines. Figure 4.1 present the results for the fixed number of assets $k = 20$ and for several sample sizes $n \in \{52, 78, 104, 130\}$, while Figure 4.2 shows similar results for fixed $n = 130$ and varying $k \in \{5, 10, 20, 50\}$. All contours are centered in the positive region of R, indicating that the return of all portfolios is positive on average. Figure 4.1 also indicates that the credible sets become smaller for larger sample sizes and are, therefore, in line with statistical theory. This indicates not only more stable expected return estimates but also more stable variance estimations. It is noteworthy that the asymptotic approximation obtained under the diffuse prior works well, while we get considerable differences for the conjugate prior. The situation is improved in Figure 4.2 when n is considerably larger than k. These findings are in line with Bayesian theory, i.e., if the value of k is comparable to n, then the impact of the prior becomes larger which influences the asymptotic approximation. Interestingly, Figure 4.2 also shows almost equal or even slightly smaller ellipses for larger portfolios. This is the result of two counteracting effects: while statistical theory implies that the uncertainty increases with the number of parameters, economic theory implies that a higher number of assets leads to a diversification effect, i.e. to the reduction of uncertainty.

Figures 4.3 and 4.4 present the 95% credible intervals for the variances of the optimal portfolios with the expected return fixed to some level R_0. The solid black line shows the Bayesian estimate of the efficient frontier, while the dashed lines in each plot are the credible sets obtained for each fixed value R_0. Additionally, we also draw two blue lines which consist of the lower and the upper limits of the credible intervals calculated for the considered values of R_0. In Figures 4.3 and 4.4 the results are shown for the diffuse prior only, while similar figures for the conjugate prior are available from the authors on request. Finally, we fix $k = 20$ and consider several $n \in \{52, 78, 104, 130\}$ in Figure 4.3, while Figure 4.4 shows the results for $n = 130$ and varying $k \in \{5, 10, 20, 50\}$.

It is interesting to observe that the estimator for the slope parameter of the efficient frontier becomes larger for smaller values of n fixed k as well as for larger k with fixed n. The values of the estimated slope parameter have strong influence on the length of credible intervals. For the same value of R_0 wider credible intervals are present in the case of smaller value of the estimated slope parameter. To this end, as in Figures 4.1 and 4.2 we do not find heavily inflated credible intervals for growing portfolio sizes, again reflecting the two counteracting effects of diversification and the number of estimated parameters.

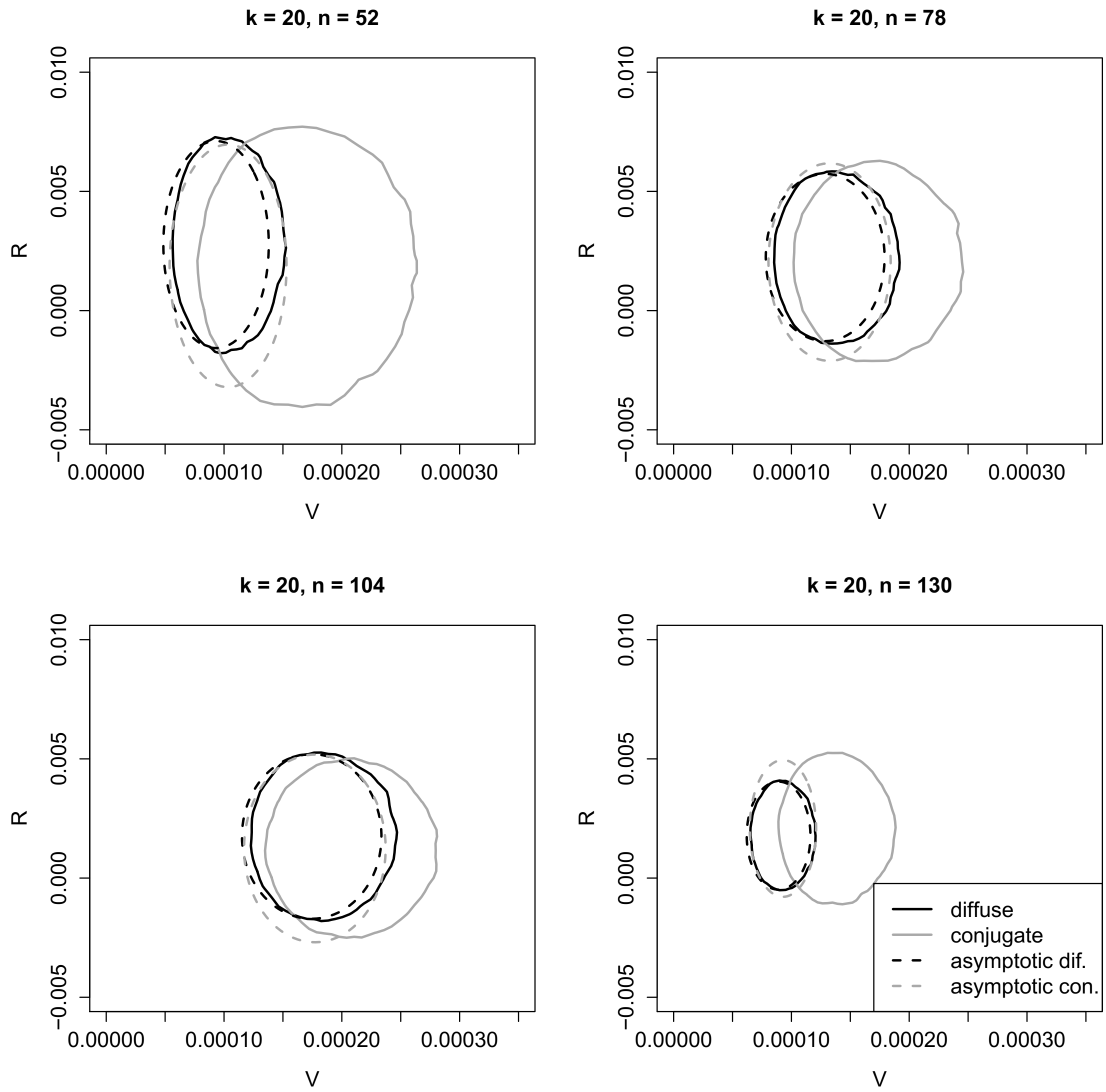

Figure 4.1: 95% credible regions for the expected return and the variance of the global minimum variance portfolio under the diffuse and conjugate priors.
We put $k = 20$ and $n \in \{52, 78, 104, 130\}$.

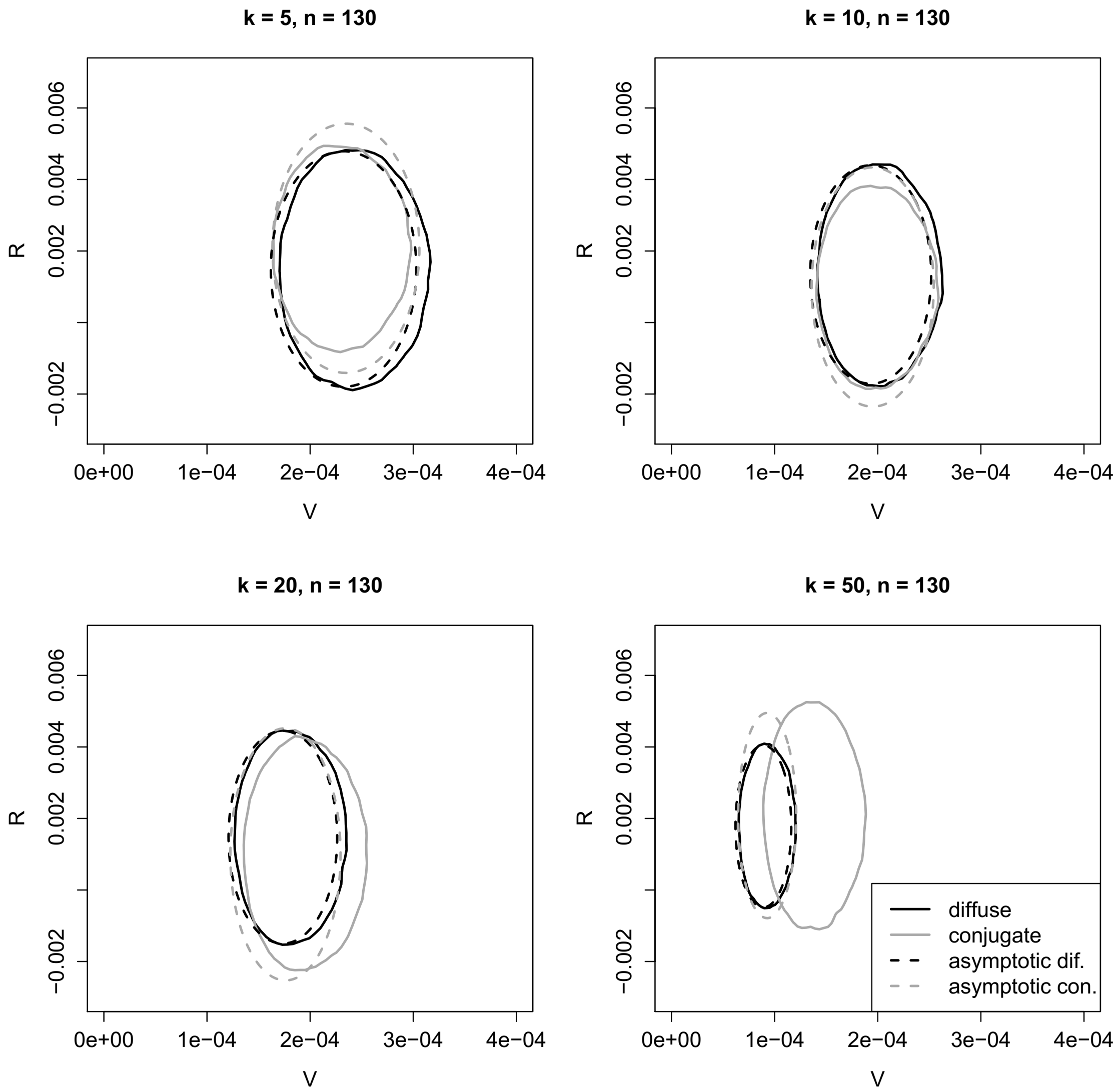

Figure 4.2: 95% credible regions for the expected return and the variance of the global minimum variance portfolio under the diffuse and conjugate priors.
We put $n = 130$ and $k \in \{5, 10, 20, 50\}$.

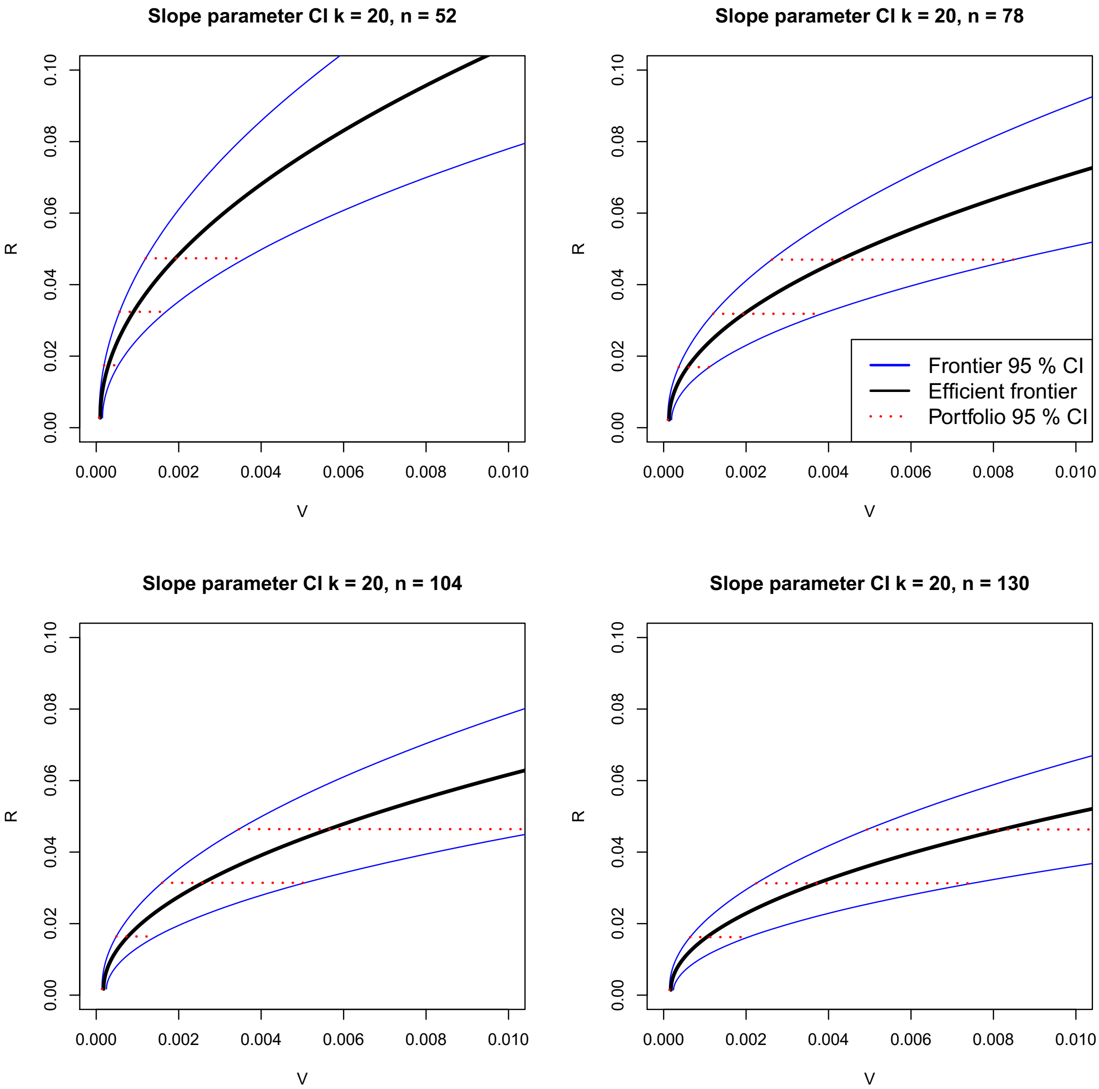

Figure 4.3: Efficient frontier for different sample sizes.
Bayes estimation of the efficient frontier (black line) together with lower and the upper limits (blue lines) of 95% credible intervals for the variance of optimal portfolios calculated for the fixed value of the expected return $R \in [0, 0.1]$. The credible intervals are red dash lines for several selected values of R. We put $k = 20$ and $n \in \{52, 78, 104, 130\}$.

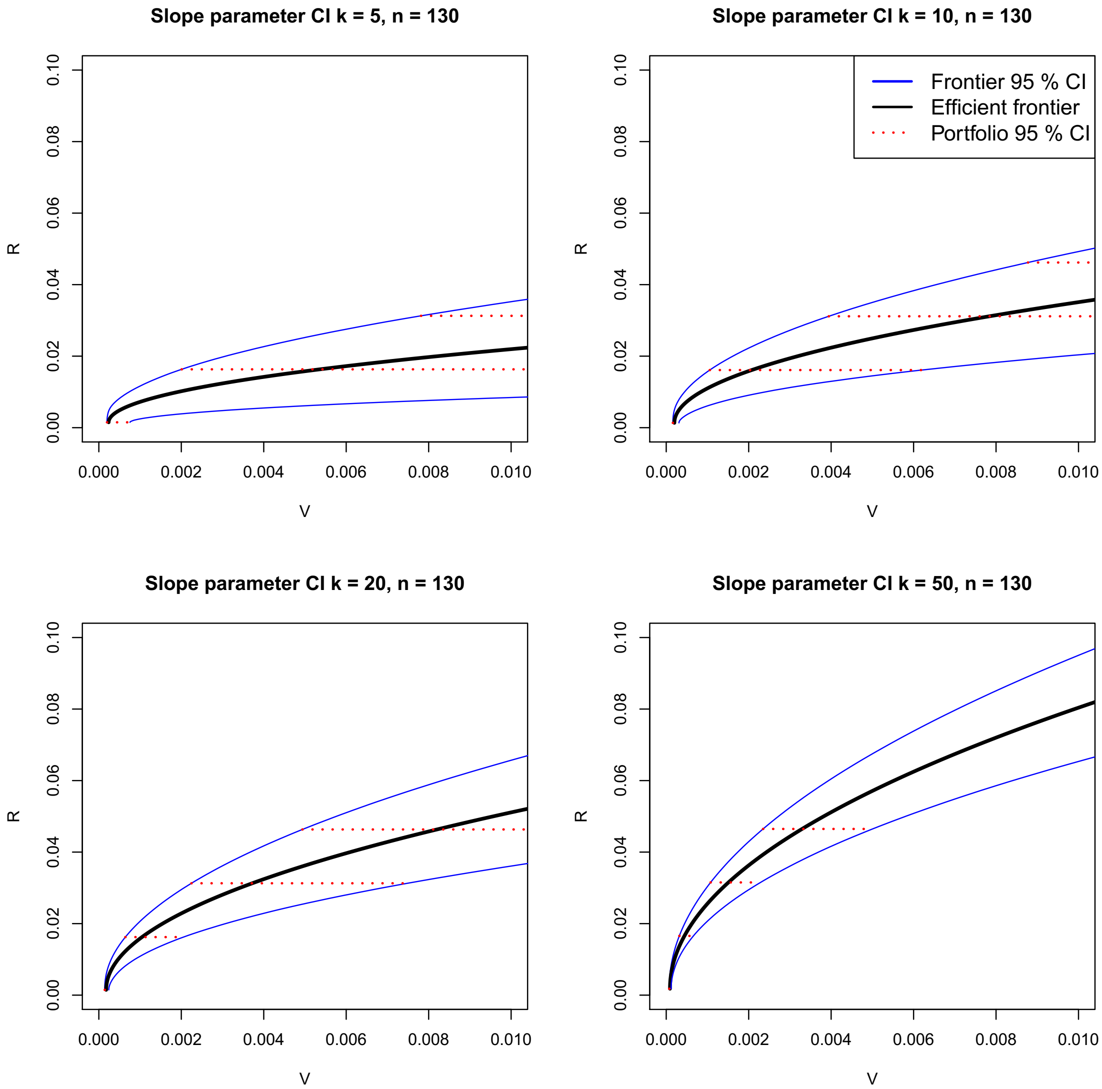

Figure 4.4: Efficient frontier for different portfolio dimensions.
Bayes estimation of the efficient frontier (black line) together with lower and the upper limits (blue lines) of 95% credible intervals for the variance of optimal portfolios calculated for the fixed value of the expected return $R \in [0, 0.1]$. The credible intervals are red dash lines for several selected values of R. We put $n = 130$ and $n \in \{5, 10, 20, 50\}$.

4.4 Summary

In this chapter we derive the posterior distributions for the expected return of the global minimum variance portfolio, of its variance, and of the slope parameter of the efficient frontier in a Bayesian framework. Using the diffuse and the conjugate prior, this allows us to incorporate subjective beliefs into an investment decision. Our contribution is the derivation of stochastic representations for all the parameters determining the efficient frontier which appears to be computationally highly efficient and, in the current situation, making MCMC-procedures obsolete. In addition to this, the stochastic representations allow us also to derive Bayesian estimates. This can be done by calculating the expectation of the stochastic representation or numerically determining distribution-based estimators like the median. We then applied the derived results in an empirical study using assets from the S&P 500 by calculating the 95% credible regions for the expected return and the variance of the global minimum variance portfolio. It turned out that the true distributions from the stochastic representation are almost covered by the derived asymptotic distributions for the diffuse prior. This result might be pleasing for moderate sample sizes. Nevertheless, the asymptotic distribution deviates sometimes strongly from the true distribution in case of the conjugate prior. Therefore, the availability of the true posterior distributions is necessary in this context. Besides this finding, the true distribution is not severely affected by the number of assets in the portfolio, even for moderate sample sizes. This can be attributed to a trade off between the number of assets increasing the estimation risk, but also to the diversification effect emphasized by Markowitz (1952) which decreases the economical risk.

Hence, this research contributes in a way that it helps to understand the effect of estimation risk in portfolio theory better, especially from the Bayesian perspective. Further research questions are the impact of different parameterizations of the conjugate prior or considering other priors, of course also for different portfolio models. The trade off between the estimation risk and the diversification effect should be studied in more detail.

4.5 Proofs and Supplementary Material

4.5.1 Supplementary lemmas

First, we present several important lemmas which are used in the proof of Theorem 13.

Lemma 8. *Let*

$$\mathbf{\Omega}|\boldsymbol{\nu}, \mathbf{y} \sim \mathcal{IW}_k(k_y, \mathbf{S}_y^*(\boldsymbol{\nu})) \quad \textit{and} \quad \boldsymbol{\nu}|\mathbf{y} \sim f(\cdot|\mathbf{y}),$$

and the symbol $f(\cdot|\mathbf{y})$ stands for the posterior distribution of $\boldsymbol{\nu}$. Let $\mathbf{M}$ be a $p \times k$ matrix of constants such that rank$(\mathbf{M}) = p < k$. Then

(a)

$$\mathbf{\Xi} = (\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top)^{-1/2}(\mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}^\top)(\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top)^{-1/2} \sim \mathcal{W}_p\left(k_y - k - 1, \mathbf{I}_p\right),$$

and it is independent of $\boldsymbol{\nu}$ and $\mathbf{y}$;

(b)

$$\boldsymbol{\eta} = \frac{(\mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}^\top)^{1/2}\left((\mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}^\top)^{-1}\mathbf{M}\mathbf{\Omega}^{-1}\boldsymbol{\nu} - (\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top)^{-1}\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}\right)}{\sqrt{\boldsymbol{\nu}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \boldsymbol{\nu}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top(\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top)^{-1}\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}}} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p),$$

and it is independent of $\boldsymbol{\nu}$ and $\mathbf{y}$;

(c)

$$\xi = \frac{\boldsymbol{\nu}^\top\mathbf{\Omega}^{-1}\boldsymbol{\nu} - \boldsymbol{\nu}^\top\mathbf{\Omega}^{-1}\mathbf{M}^\top(\mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}^\top)^{-1}\mathbf{M}\mathbf{\Omega}^{-1}\boldsymbol{\nu}}{\boldsymbol{\nu}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \boldsymbol{\nu}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top(\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{M}^\top)^{-1}\mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}} \sim \chi^2_{k_y-k-p-1},$$

and it is independent of $\boldsymbol{\nu}$ and $\mathbf{y}$;

(d) ξ, $\boldsymbol{\eta}$, and $\mathbf{\Xi}$ are independently distributed.

Proof. The application Theorem 3.4.1 of Gupta and Nagar (2000) leads to

$$\mathbf{\Omega}^{-1}|\boldsymbol{\nu}, \mathbf{y} \sim \mathcal{W}_k(k_y - k - 1, \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}).$$

Fix $\boldsymbol{\nu} = \boldsymbol{\nu}^*$ and define $\tilde{\mathbf{M}} = (\mathbf{M}^\top, \boldsymbol{\nu}^*)^\top$. Let

$$\hat{\mathbf{H}} = \tilde{\mathbf{M}}\mathbf{\Omega}^{-1}\tilde{\mathbf{M}}^\top = \begin{pmatrix} \mathbf{M}\mathbf{\Omega}^{-1}\mathbf{M}^\top & \mathbf{M}\mathbf{\Omega}^{-1}\boldsymbol{\nu}^* \\ \boldsymbol{\nu}^{*\top}\mathbf{\Omega}^{-1}\mathbf{M}^\top & \boldsymbol{\nu}^{*\top}\mathbf{\Omega}^{-1}\boldsymbol{\nu}^* \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{H}}_{11} & \hat{\mathbf{H}}_{12} \\ \hat{\mathbf{H}}_{21} & \hat{\mathbf{H}}_{22} \end{pmatrix}$$

with $\hat{\mathbf{H}}_{22} = \boldsymbol{\nu}^{*\top}\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}^*$ and

$$\mathbf{H} = \tilde{\mathbf{M}}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\tilde{\mathbf{M}}^\top = \begin{pmatrix} \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\mathbf{M}^\top & \mathbf{M}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\boldsymbol{\nu}^* \\ \boldsymbol{\nu}^{*\top}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\mathbf{M}^\top & \boldsymbol{\nu}^{*\top}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\boldsymbol{\nu}^* \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}$$

with $\mathbf{H}_{22} = \boldsymbol{\nu}^{*\top}\mathbf{S}_y^*(\boldsymbol{\nu}^*)^{-1}\boldsymbol{\nu}^*$.

From Theorem 3.2.5 in Muirhead (1982) we get that $\hat{\mathbf{H}}|\mathbf{y} \sim \mathcal{W}_{p+1}(k_y - k - 1, \mathbf{H})$. Moreover, from Theorem 3.2.10 of Muirhead (1982) we obtain that

(i) $\hat{\mathbf{H}}_{11}|\mathbf{y} \sim \mathcal{W}_p(k_y - k - 1, \mathbf{H}_{11})$;

(ii) $\hat{\mathbf{H}}_{11}^{-1}\hat{\mathbf{H}}_{12}|\hat{\mathbf{H}}_{11}, \mathbf{y} \sim \mathcal{N}_p(\mathbf{H}_{11}^{-1}\mathbf{H}_{12}, \mathbf{H}_{22\cdot1}\hat{\mathbf{H}}_{11}^{-1})$ where $\mathbf{H}_{22\cdot1} = \mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^{-1}\mathbf{H}_{12}$;

(iii) $\hat{\mathbf{H}}_{22\cdot1} = \hat{\mathbf{H}}_{22} - \hat{\mathbf{H}}_{21}\hat{\mathbf{H}}_{11}^{-1}\hat{\mathbf{H}}_{12}|\mathbf{y} \sim \mathcal{W}_1(k_y - k - p - 1, \mathbf{H}_{22\cdot1})$ and it is independent of $\hat{\mathbf{H}}_{11}$ and $\hat{\mathbf{H}}_{12}$.

The rest of the proof follows from the standardization of the multivariate normal distribution and the application of Theorems 3.2.5 and 3.2.8 in Muirhead (1982). □

Lemma 9. *Let*

$$\boldsymbol{\Omega}|\boldsymbol{\nu}, \mathbf{y} \sim \mathcal{IW}_k(k_y, \mathbf{S}_y^*(\boldsymbol{\nu})) \quad \text{and} \quad \boldsymbol{\nu}|\mathbf{y} \sim f(\cdot|\mathbf{y}),$$

and the symbol $f(\cdot|\mathbf{y})$ stands for the posterior distribution of $\boldsymbol{\nu}$. Then

(a)

$$\frac{1}{\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1}} \stackrel{d}{=} \frac{1}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}\tau^{-1}$$

(b)

$$\frac{\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1}} \stackrel{d}{=} \frac{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}} + \frac{\eta}{\sqrt{\tau}}\sqrt{\frac{\boldsymbol{\nu}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}}$$

(c)

$$\boldsymbol{\nu}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1}} \stackrel{d}{=} \xi\left(\boldsymbol{\nu}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}\right)$$

(d) $\xi \sim \chi^2_{k_y-k-2}$, $\eta \sim \mathcal{N}(0,1)$, $\tau \sim \chi^2_{k_y-k-1}$ and they are mutually independently distributed.

Proof. It holds that

$$\frac{1}{\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1}} = \frac{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}{\mathbf{1}^\top\boldsymbol{\Omega}^{-1}\mathbf{1}}\frac{1}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}} \stackrel{d}{=} \frac{1}{\mathbf{1}^\top\mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}\tau^{-1},$$

$$
\begin{aligned}
\frac{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}} &= \sqrt{\frac{\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}}{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}}} \sqrt{\frac{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}}{\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}}} \\
&\times \left(\frac{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}} - \frac{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}\right) + \frac{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}} \\
&\stackrel{d}{=} \frac{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu}}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}} + \frac{\eta}{\sqrt{\tau}} \sqrt{\frac{\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}},
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\nu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}} &= \frac{\boldsymbol{\nu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}}}{\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}} \left(\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}\right) \\
&\stackrel{d}{=} \xi \left(\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} - \frac{(\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu})^2}{\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1}}\right),
\end{aligned}
$$

where the equalities in distribution follow from Lemma 8. Moreover, from that lemma we also get that ξ, η, τ are mutually independent and they are independent from $\boldsymbol{\nu}$. $\square$

Lemma 10. *Under assumption of Lemma 9, let*

$$
\boldsymbol{\nu}|\mathbf{y} \sim t_k\left(d_y, \mathbf{m}_y, \mathbf{S}_y/d_y\right) \quad \text{and} \quad \mathbf{S}_y^*(\boldsymbol{\nu}) = v_y(\mathbf{S}_y + (\boldsymbol{\nu} - \mathbf{m}_y)(\boldsymbol{\nu} - \mathbf{m}_y)^\top).
$$

Then

(a)

$$
\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1} \stackrel{d}{=} v_y^{-1}\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}\left(1 - \frac{\psi_1^2}{\phi + \varphi + \psi_1^2 + \psi_2^2}\right);
$$

(b)

$$
\begin{aligned}
\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} &\stackrel{d}{=} v_y^{-1}\left(\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y + \frac{\sqrt{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}}\psi_1}{\phi + \varphi + \psi_1^2 + \psi_2^2}\sqrt{\phi}\right. \\
&- \left.\frac{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y\psi_1^2 + \sqrt{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}\mathbf{m}_y^\top \mathbf{S}_y^{-1}\mathbf{m}_y - (\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y)^2}\psi_1\psi_2}{\phi + \varphi + \psi_1^2 + \psi_2^2}\right);
\end{aligned}
$$

(c)

$$
\begin{aligned}
\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} &\overset{d}{=} v_y^{-1}\Bigg(\mathbf{m}_y^\top \mathbf{S}_y^{-1}\mathbf{m}_y + \frac{\varphi+\psi_1^2+\psi_2^2}{\phi+\varphi+\psi_1^2+\psi_2^2}\\
&+ 2\frac{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y\psi_1 + \sqrt{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}\mathbf{m}_y^\top \mathbf{S}_y^{-1}\mathbf{m}_y - (\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y)^2}\psi_2}{\phi+\varphi+\psi_1^2+\psi_2^2}\frac{\sqrt{\phi}}{\sqrt{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}}}\\
&- \frac{(\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y\psi_1 + \sqrt{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}\mathbf{m}_y^\top \mathbf{S}_y^{-1}\mathbf{m}_y - (\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y)^2}\psi_2)^2}{\phi+\varphi+\psi_1^2+\psi_2^2}\frac{1}{\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1}}\Bigg);
\end{aligned}
$$

(d) $\varphi \sim \chi^2_{k-2}$, $\phi \sim \chi^2_{d_y}$, $\boldsymbol{\psi} \sim \mathcal{N}_2(\mathbf{0},\mathbf{I})$, *and they are independent*

Proof. The application of the Sherman-Morrison formula (see, e.g., Meyer (2000, p.125)) yields

$$
(\mathbf{S}_y + (\boldsymbol{\nu}-\mathbf{m}_y)(\boldsymbol{\nu}-\mathbf{m}_y)^\top)^{-1} = \mathbf{S}_y^{-1} - \frac{\mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}
$$

Then, we get

$$
\begin{aligned}
\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\mathbf{1} &= v_y^{-1}\left(\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{1} - \frac{\left(\mathbf{1}^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)\right)^2}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}\right), \quad (4.26)\\
\mathbf{1}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} &= v_y^{-1}\Bigg(\mathbf{1}^\top \mathbf{S}_y^{-1}\mathbf{m}_y - \frac{\mathbf{1}^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)\mathbf{m}_y^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)} \quad (4.27)\\
&+ \frac{\mathbf{1}^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}\Bigg),\\
\boldsymbol{\nu}^\top \mathbf{S}_y^*(\boldsymbol{\nu})^{-1}\boldsymbol{\nu} &= v_y^{-1}\Bigg(\mathbf{m}_y^\top \mathbf{S}_y^{-1}\mathbf{m}_y + \frac{(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)} \quad (4.28)\\
&+ 2\frac{\mathbf{m}_y^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)} - \frac{(\mathbf{m}_y^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y))^2}{1+(\boldsymbol{\nu}-\mathbf{m}_y)^\top \mathbf{S}_y^{-1}(\boldsymbol{\nu}-\mathbf{m}_y)}\Bigg).
\end{aligned}
$$

Using that $\sqrt{d_y}\mathbf{S}_y^{-1/2}(\boldsymbol{\nu}-\mathbf{m}_y) \sim t_k\left(d_y,\mathbf{0},\mathbf{I}\right)$ and the properties of multivariate t-distribution, we get

$$
\mathbf{S}_y^{-1/2}(\boldsymbol{\nu}-\mathbf{m}_y) \overset{d}{=} \frac{\boldsymbol{\zeta}}{\sqrt{\phi}},
$$

where $\boldsymbol{\zeta} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$ and $\phi \sim \chi^2_{d_y}$. Let

$$\mathbf{a}_1 = \mathbf{S}_y^{-1/2}\mathbf{1}, \mathbf{a}_2 = \mathbf{S}_y^{-1/2}\mathbf{m}_y, \mathbf{A} = \begin{pmatrix} \mathbf{a}_1^\top \mathbf{a}_1 & \mathbf{a}_1^\top \mathbf{a}_2 \\ \mathbf{a}_1^\top \mathbf{a}_2 & \mathbf{a}_2^\top \mathbf{a}_2 \end{pmatrix}$$

and define $\mathbf{B} = [\mathbf{a}_1\ \mathbf{a}_2]\mathbf{A}^{-1}[\mathbf{a}_1\ \mathbf{a}_2]^\top$ Then, it holds that

$$\boldsymbol{\zeta}^\top \boldsymbol{\zeta} \quad = \quad \boldsymbol{\zeta}^\top(\mathbf{I} - \mathbf{B})\boldsymbol{\zeta} + \boldsymbol{\zeta}^\top \mathbf{B}\boldsymbol{\zeta} \stackrel{d}{=} \varphi + \boldsymbol{\psi}^\top \boldsymbol{\psi},$$

where

$$\boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \mathbf{A}^{-1/2} \begin{pmatrix} \mathbf{a}_1^\top \boldsymbol{\zeta} \\ \mathbf{a}_2^\top \boldsymbol{\zeta} \end{pmatrix},$$

from which in using the Cholesky decomposition of $\mathbf{A}$ we also obtain

$$\begin{aligned} \mathbf{a}_1^\top \boldsymbol{\zeta} &= \sqrt{\mathbf{a}_1^\top \mathbf{a}_1}\psi_1, \\ \mathbf{a}_2^\top \boldsymbol{\zeta} &= \frac{\mathbf{a}_1^\top \mathbf{a}_2}{\sqrt{\mathbf{a}_1^\top \mathbf{a}_1}}\psi_1 + \sqrt{\mathbf{a}_2^\top \mathbf{a}_2 - \frac{(\mathbf{a}_1^\top \mathbf{a}_2)^2}{\mathbf{a}_1^\top \mathbf{a}_1}}\psi_2. \end{aligned}$$

The application of $\boldsymbol{\zeta} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, $\mathbf{B}$ is an idempotent matrix, and $(\mathbf{I} - \mathbf{B})\mathbf{B} = 0$ leads to the fact that φ and $\boldsymbol{\psi}$ are independent. Furthermore, from the definition of $\boldsymbol{\psi}$ we get that $\boldsymbol{\psi} \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I})$. Finally, the equalities $(\mathbf{I} - \mathbf{B})^2 = \mathbf{I} - \mathbf{B}$ and $\mathrm{tr}(\mathbf{I} - \mathbf{B}) = k - 2$ ensure that $\varphi \sim \chi^2_{k-2}$ (c.f., Mathai and Provost (1992, Theorem 5.1.1)). The application of these results together with (4.26)-(4.28) leads to the statement of the lemma. □

Lemma 11. *Let $\zeta_1 \sim \chi^2_{d_1}$ and $\zeta_2 \sim \chi^2_{d_2}$ be two independent random variables. Then $\zeta_1/(\zeta_1 + \zeta_2)$ is independent of $\zeta_1 + \zeta_2$.*

Proof of Lemma 11. Since $\zeta_1 \sim \chi^2_{d_1}$ and $\zeta_2 \sim \chi^2_{d_2}$, there exist $\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \sim \mathcal{N}_{d_1+d_2}(\mathbf{0}, \mathbf{I})$ with $\mathbf{Z}_1 : d_1 \times 1$ and $\mathbf{Z}_2 : d_2 \times 1$ such that $\zeta_1 = \mathbf{Z}_1^\top \mathbf{Z}_1$ and $\zeta_2 = \mathbf{Z}_2^\top \mathbf{Z}_2$. Since $\mathbf{Z}$ is standard normally distributed and, hence, it belongs to the class of elliptical distributions we get that $\mathbf{Z}/\sqrt{\mathbf{Z}^\top \mathbf{Z}}$ is independent of $\mathbf{Z}^\top \mathbf{Z}$ and, consequently, $\mathbf{Z}_1/\sqrt{\mathbf{Z}^\top \mathbf{Z}}$ is independent of $\mathbf{Z}^\top \mathbf{Z}$ which leads to the independence $\mathbf{Z}_1^\top \mathbf{Z}_1/\mathbf{Z}^\top \mathbf{Z}$ of $\mathbf{Z}^\top \mathbf{Z}$. Thus, $\zeta_1/(\zeta_1 + \zeta_2)$ and $\zeta_1 + \zeta_2$ are independent. □

4.5.2 Proof of theorems

Proof of Theorem 13. The results of Theorem 13 follow from Lemma 9 and 10 with $\boldsymbol{\Sigma} = \boldsymbol{\Omega}$ and $\boldsymbol{\nu} = \boldsymbol{\mu}$.

(a) From Lemma 9 we get with $k_y = n+k+1$ the following result in the case of the diffuse prior

$$V_{GMV} \stackrel{d}{=} \frac{1}{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\mathbf{1}}\tau^{-1},$$

$$R_{GMV} \stackrel{d}{=} \frac{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\mathbf{1}} + \frac{\eta}{\sqrt{\tau}}\sqrt{\frac{\boldsymbol{\mu}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu})^2}{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\mathbf{1}}}{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\mathbf{1}}},$$

$$s \stackrel{d}{=} \xi\left(\boldsymbol{\mu}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu})^2}{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\mathbf{1}}\right)$$

where the application of Lemma 10 with $d_y = n-k$, $v_y = n$, $\mathbf{m}_y = \overline{\mathbf{x}}_d$, $\mathbf{S}_y = \mathbf{S}_d/n$, and $\mathbf{S}_d^*(\boldsymbol{\nu}) = \mathbf{S}_d^*(\boldsymbol{\mu})$ leads to

$$\frac{1}{\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\mathbf{1}} \stackrel{d}{=} \frac{1}{\mathbf{1}^\top \mathbf{S}_d^{-1}\mathbf{1}}\frac{\phi+\varphi+\psi_1^2+\psi_2^2}{\phi+\varphi+\psi_2^2} = V_{GMV;d}\left(1+\frac{\psi_1^2}{\phi+\varphi+\psi_2^2}\right),$$

$$\begin{aligned}\mathbf{1}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} &\stackrel{d}{=} \frac{R_{GMV;d}}{V_{GMV;d}} - \frac{\frac{R_{GMV;d}}{V_{GMV;d}}\psi_1^2 + \sqrt{\frac{s_d}{V_{GMV;d}}}\psi_1\psi_2}{\phi+\varphi+\psi_1^2+\psi_2^2}\\ &+ \frac{\psi_1}{\phi+\varphi+\psi_1^2+\psi_2^2}\frac{\sqrt{\phi}}{\sqrt{n}\sqrt{V_{GMV;d}}},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}^\top \mathbf{S}_d^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} &\stackrel{d}{=} s_d + \frac{R_{GMV;d}^2}{V_{GMV;d}} + \frac{1}{n}\frac{\varphi+\psi_1^2+\psi_2^2}{\phi+\varphi+\psi_1^2+\psi_2^2}\\ &+ 2\frac{\frac{R_{GMV;d}}{V_{GMV;d}}\psi_1 + \sqrt{\frac{s_d}{V_{GMV;d}}}\psi_2}{\phi+\varphi+\psi_1^2+\psi_2^2}\frac{\sqrt{V_{GMV;d}}\sqrt{\phi}}{\sqrt{n}} - \frac{\left(\frac{R_{GMV;d}}{V_{GMV;d}}\psi_1 + \sqrt{\frac{s_d}{V_{GMV;d}}}\psi_2\right)^2}{\phi+\varphi+\psi_1^2+\psi_2^2}V_{GMV;d}\end{aligned}$$

with $\xi \sim \chi^2_{n-1}$, $\tau \sim \chi^2_n$, $\varphi \sim \chi^2_{k-2}$, $\phi \sim \chi^2_{n-k}$, $\eta \sim \mathcal{N}(0,1)$, and $\boldsymbol{\psi} \sim \mathcal{N}_2(\mathbf{0},\mathbf{I})$, and they are mutually independently distributed.

Putting these results together we get the statement of Theorem 13.(a).

(b) From Lemma 9 we get with $k_y = n+d_0+1$ the following result in the case of the conjugate prior

$$V_{GMV} \stackrel{d}{=} \frac{1}{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\mathbf{1}}\tau^{-1},$$

$$R_{GMV} \stackrel{d}{=} \frac{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\mathbf{1}} + \frac{\eta}{\sqrt{\tau}}\sqrt{\frac{\boldsymbol{\mu}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu})^2}{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\mathbf{1}}}{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\mathbf{1}}},$$

$$s \stackrel{d}{=} \xi \left(\boldsymbol{\mu}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu})^2}{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\mathbf{1}} \right)$$

where the application of Lemma 10 with $d_y = n + d_0 - 2k$, $v_y = n + r_0$, $\mathbf{m}_y = \overline{\mathbf{x}}_c$, $\mathbf{S}_y = \mathbf{S}_c/(n + r_0)$, and $\mathbf{S}_y^*(\boldsymbol{\nu}) = \mathbf{S}_c^*(\boldsymbol{\mu})$ leads to

$$\frac{1}{\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\mathbf{1}} \stackrel{d}{=} \frac{1}{\mathbf{1}^\top \mathbf{S}_c^{-1}\mathbf{1}} \frac{\phi + \varphi + \psi_1^2 + \psi_2^2}{\phi + \varphi + \psi_2^2} = V_{GMV;c}\left(1 + \frac{\psi_1^2}{\phi + \varphi + \psi_2^2}\right),$$

$$\begin{aligned}\mathbf{1}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} &\stackrel{d}{=} \frac{R_{GMV;c}}{V_{GMV;c}} - \frac{\frac{R_{GMV;c}}{V_{GMV;c}}\psi_1^2 + \sqrt{\frac{s_c}{V_{GMV;c}}}\psi_1\psi_2}{\phi + \varphi + \psi_1^2 + \psi_2^2} \\ &+ \frac{\psi_1}{\phi + \varphi + \psi_1^2 + \psi_2^2}\frac{\sqrt{\phi}}{\sqrt{n + r_0}\sqrt{V_{GMV;c}}},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}^\top \mathbf{S}_c^*(\boldsymbol{\mu})^{-1}\boldsymbol{\mu} &\stackrel{d}{=} s_c + \frac{R_{GMV;c}^2}{V_{GMV;c}} + \frac{1}{n + r_0}\frac{\varphi + \psi_1^2 + \psi_2^2}{\phi + \varphi + \psi_1^2 + \psi_2^2} \\ &+ 2\frac{\frac{R_{GMV;c}}{V_{GMV;c}}\psi_1 + \sqrt{\frac{s_c}{V_{GMV;c}}}\psi_2}{\phi + \varphi + \psi_1^2 + \psi_2^2}\frac{\sqrt{V_{GMV;c}}\sqrt{\phi}}{\sqrt{n + r_0}} - \frac{\left(\frac{R_{GMV;c}}{V_{GMV;c}}\psi_1 + \sqrt{\frac{s_c}{V_{GMV;c}}}\psi_2\right)^2}{\phi + \varphi + \psi_1^2 + \psi_2^2}V_{GMV;c}\end{aligned}$$

with $\xi \sim \chi^2_{n+d_0-k-1}$, $\tau \sim \chi^2_{n+d_0-k}$, $\varphi \sim \chi^2_{k-2}$, $\phi \sim \chi^2_{n+d_0-2k}$, $\eta \sim \mathcal{N}(0, 1)$, and $\boldsymbol{\psi} \sim \mathcal{N}_2(\mathbf{0}, \mathbf{I})$, and they are mutually independently distributed. This completes the proof of Theorem 13.

□

Proof of Theorem 14. (a) Using the stochastic representation of Theorem 13 and the distributional properties of τ, ψ_1, ϕ, φ and ψ_2 we get

$$\begin{aligned}\hat{V}_{GMV;d} &= \mathbb{E}(V_{GMV}|\mathbf{x}) = V_{GMV;d}\left(1 + \mathbb{E}(\psi_1^2)\mathbb{E}\left(\frac{1}{\phi + \varphi + \psi_2^2}\right)\right)\mathbb{E}(\tau^{-1}) \\ &= V_{GMV;d}\left(1 + \frac{1}{n-3}\right)\frac{1}{n-2} = \frac{1}{n-3}V_{GMV;d}\end{aligned}$$

Similarly, in using that $\mathbb{E}(\psi_1) = \mathbb{E}(\eta) = 0$ and the independence between the random variables in (4.11), we obtain

$$\hat{R}_{GMV;d} = \mathbb{E}(R_{GMV}|\mathbf{x}) = R_{GMV;d}$$

Finally, it holds that

$$\begin{aligned}\hat{s}_d &= \mathbb{E}(s|\mathbf{x}) = \mathbb{E}(\xi)\Bigg(s_d + \frac{1}{n} - s_d\mathbb{E}\left(\frac{\psi_2^2}{\phi+\varphi+\psi_2^2}\right) \\ &+ 2\frac{\sqrt{s_d}}{\sqrt{n}}\mathbb{E}\left(\frac{\psi_2\sqrt{\phi}}{\phi+\varphi+\psi_2^2}\right) - \frac{1}{n}\mathbb{E}\left(\frac{\phi}{\phi+\varphi+\psi_2^2}\right)\Bigg).\end{aligned}$$

First, we note that $\frac{\psi_2\sqrt{\phi}}{\phi+\varphi+\psi_2^2}$ is a symmetrically around zero distributed random variable and, hence, its expectation is zero. Second, since ϕ, φ, and ψ_2 are independent χ^2-distributed random variables, we get that they are also independent gamma-distributed random variables with $\phi \sim \text{Gamma}((n-k)/2, 2)$, $\varphi \sim \text{Gamma}((k-2)/2, 2)$, and $\psi_2 \sim \text{Gamma}(1/2, 2)$ and consequently, the random vector $\left(\frac{\phi}{\phi+\varphi+\psi_2^2}, \frac{\varphi}{\phi+\varphi+\psi_2^2}, \frac{\psi_2^2}{\phi+\varphi+\psi_2^2}\right)^\top$ has a Dirichlet distribution with parameter vector $((n-k)/2,\ (k-2)/2, 1/2)^\top$. Hence,

$$\mathbb{E}\left(\frac{\psi_2^2}{\phi+\varphi+\psi_2^2}\right) = \frac{1}{n-1} \quad \text{and} \quad \mathbb{E}\left(\frac{\phi}{\phi+\varphi+\psi_2^2}\right) = \frac{n-k}{n-1},$$

which leads to

$$\hat{s}_d = (n-1)\left(s_d + \frac{1}{n} - s_d\frac{1}{n-1} - \frac{1}{n}\frac{n-k}{n-1}\right) = (n-2)s_d + \frac{k-1}{n}.$$

(b) Similarly, we get

$$\begin{aligned}\hat{V}_{GMV;c} &= \mathbb{E}(V_{GMV}|\mathbf{x}) = V_{GMV;c}\left(1 + \frac{1}{n+d_0-k-3}\right)\frac{1}{n+d_0-k-2} \\ &= \frac{1}{n+d_0-k-3}V_{GMV;c}, \\ \hat{R}_{GMV;c} &= \mathbb{E}(R_{GMV}|\mathbf{x}) = R_{GMV;c}, \\ \hat{s}_c &= \mathbb{E}(s|\mathbf{x}) = (n+d_0-k-1)\Bigg(s_c + \frac{1}{n+r_0} - \frac{s_c}{n+d_0-k-1} \\ &- \frac{n+d_0-2k}{(n+r_0)(n+d_0-k-1)}\Bigg) = (n+d_0-k-2)s_c + \frac{k-1}{n+r_0}.\end{aligned}$$

□

Proof of Theorem 15. From the discussion presented after the proof of Theorem 13 we know that V_{GMV} and s are independently distributed under both priors. The equalities $\mathbb{C}ov(R_{GMV}, V_{GMV}|\mathbf{x}) = \mathbb{C}ov(R_{GMV}, s|\mathbf{x}) = 0$ follows directly from Theorem 13 and the equalities $\mathbb{E}(\psi_1) = \mathbb{E}(\eta) = 0$ under both priors. Next, we derive the expressions for the variances.

(a) Using the stochastic representation of Theorem 13 and the distributional properties of τ, ψ_1, ϕ, φ and ψ_2, we get

$$
\begin{aligned}
\mathbb{V}ar(V_{GMV}|\mathbf{x}) &= V_{GMV;d}^2\left(1+2\mathbb{E}(\psi_1^2)\mathbb{E}\left(\frac{1}{\phi+\varphi+\psi_2^2}\right)+\mathbb{E}(\psi_1^4)\mathbb{E}\left(\frac{1}{(\phi+\varphi+\psi_2^2)^2}\right)\right)\\
\times&\quad \mathbb{E}(\tau^{-2})-\mathbb{E}(V_{GMV}|\mathbf{x})^2\\
=&\quad V_{GMV;d}^2\left(1+\frac{2}{n-3}+\frac{3}{(n-3)(n-5)}\right)\frac{1}{(n-2)(n-4)}-V_{GMV;d}^2\frac{1}{(n-3)^2}\\
=&\quad 2V_{GMV;d}^2\frac{1}{(n-3)^2(n-5)}\,.
\end{aligned}
$$

Similarly, in using that $\mathbb{E}(\psi_1)=\mathbb{E}(\eta)=0$ and $\mathbb{E}(\psi_1^2)=\mathbb{E}(\eta^2)=1$, Lemma 11 which implies that $\phi+\varphi+\psi_2^2$ and $(\sqrt{s_d}\psi_2-\sqrt{\phi/n})/\sqrt{\phi+\varphi+\psi_2^2}$ are independent, as well as the independence between the random variables in (4.11), we obtain

$$
\begin{aligned}
\mathbb{V}ar(R_{GMV}|\mathbf{x}) &= V_{GMV;d}\mathbb{E}\left(\frac{(\sqrt{s_d}\psi_2-\sqrt{\phi/n})^2}{\phi+\varphi+\psi_2^2}\right)\mathbb{E}\left(\frac{1}{\phi+\varphi+\psi_2^2}\right)\\
+&\quad V_{GMV;d}\mathbb{E}(\tau^{-1})\mathbb{E}\left(1+\frac{\psi_1^2}{\phi+\varphi+\psi_2^2}\right)\mathbb{E}\left(s_d+\frac{1}{n}-\frac{(\sqrt{s_d}\psi_2-\sqrt{\phi/n})^2}{\phi+\varphi+\psi_2^2}\right)\\
=&\quad V_{GMV;d}\left(s_d\frac{1}{n-1}+\frac{n-k}{n(n-1)}\right)\frac{1}{n-3}+V_{GMV;d}\frac{1}{n-3}\left(\frac{n-2}{n-1}s_d+\frac{k-1}{n(n-1)}\right)\\
=&\quad V_{GMV;d}\left(s_d+\frac{1}{n}\right)\frac{1}{n-3},
\end{aligned}
$$

where the first equality follows from the proof of Theorem 14.

Finally, it holds that

$$\begin{aligned}
\mathbb{E}(s^2|\mathbf{x}) &= \mathbb{E}(\xi^2)\Bigg(\left(s_d+\frac{1}{n}\right)^2+\mathbb{E}\left(\frac{(\sqrt{s_d}\psi_2-\sqrt{\phi/n})^4}{(\phi+\varphi+\psi_2^2)^2}\right)\\
&- 2\left(s_d+\frac{1}{n}\right)\mathbb{E}\left(\frac{(\sqrt{s_d}\psi_2-\sqrt{\phi/n})^2}{\phi+\varphi+\psi_2^2}\right)\Bigg)\\
&= (n^2-1)\Bigg(\left(s_d+\frac{1}{n}\right)^2+s_d^2\mathbb{E}\left(\frac{\psi_2^4}{(\phi+\varphi+\psi_2^2)^2}\right)\\
&+ 6\frac{s_d}{n}\mathbb{E}\left(\frac{\psi_2^2\phi}{(\phi+\varphi+\psi_2^2)^2}\right)+\frac{1}{n^2}\mathbb{E}\left(\frac{\phi^2}{(\phi+\varphi+\psi_2^2)^2}\right)\\
&- 2\left(s_d+\frac{1}{n}\right)\left(s_d\mathbb{E}\left(\frac{\psi_2^2}{\phi+\varphi+\psi_2^2}\right)+\frac{1}{n}\mathbb{E}\left(\frac{\phi}{\phi+\varphi+\psi_2^2}\right)\right)\Bigg),
\end{aligned}$$

where we use that $\psi_2\phi^{3/2}/(\phi+\varphi+\psi_2^2)^2$, $\psi_2^3\phi^{1/2}/(\phi+\varphi+\psi_2^2)^2$, and $\psi_2\phi^{1/2}/(\phi+\varphi+\psi_2^2)$ are random variables which are symmetrically distributed around zero. The application of

$$\mathbb{E}\left(\frac{\psi_2^2}{\phi+\varphi+\psi_2^2}\right)=\frac{1}{n-1}\quad\text{and}\quad\mathbb{E}\left(\frac{\phi}{\phi+\varphi+\psi_2^2}\right)=\frac{n-k}{n-1},$$

and

$$\begin{aligned}
&\mathbb{E}\left(\frac{\psi_2^4}{(\phi+\varphi+\psi_2^2)^2}\right)=\mathbb{V}ar\left(\frac{\psi_2^2}{\phi+\varphi+\psi_2^2}\right)+\frac{1}{(n-1)^2}\\
=\ &2\frac{n-2}{(n+1)(n-1)^2}+\frac{1}{(n-1)^2},\\
&\mathbb{E}\left(\frac{\psi_2^2\phi}{(\phi+\varphi+\psi_2^2)^2}\right)=\mathbb{C}ov\left(\frac{\psi_2^2}{\phi+\varphi+\psi_2^2},\frac{\phi}{\phi+\varphi+\psi_2^2}\right)+\frac{n-k}{(n-1)^2}\\
=\ &-2\frac{n-k}{(n+1)(n-1)^2}+\frac{n-k}{(n-1)^2},\\
&\mathbb{E}\left(\frac{\phi^2}{(\phi+\varphi+\psi_2^2)^2}\right)=\mathbb{V}ar\left(\frac{\phi}{\phi+\varphi+\psi_2^2}\right)+\frac{(n-k)^2}{(n-1)^2}\\
=\ &2\frac{(n-k)(k-1)}{(n+1)(n-1)^2}+\frac{(n-k)^2}{(n-1)^2}.
\end{aligned}$$

leads to

$$
\begin{aligned}
\mathbb{V}ar(s|\mathbf{x}) = (n^2-1)\Bigg(&\left(s_d+\frac{1}{n}\right)^2 + s_d^2\left(2\frac{n-2}{(n+1)(n-1)^2}+\frac{1}{(n-1)^2}\right)\\
+ & 6\frac{s_d}{n}\left(-2\frac{n-k}{(n+1)(n-1)^2}+\frac{n-k}{(n-1)^2}\right)+\frac{1}{n^2}\left(2\frac{(n-k)(k-1)}{(n+1)(n-1)^2}+\frac{(n-k)^2}{(n-1)^2}\right)\\
- & 2\left(s_d+\frac{1}{n}\right)\left(s_d\frac{1}{n-1}+\frac{1}{n}\frac{n-k}{n-1}\right)\Bigg)-\left((n-2)s_d+\frac{k-1}{n}\right)^2\\
= & s_d^2\left(n^2-1+2\frac{n-2}{n-1}+\frac{n+1}{n-1}-2(n+1)-(n-2)^2\right)\\
+ & 2\frac{s_d}{n}\left(n^2-1-6\frac{n-k}{n-1}+3\frac{(n-k)(n+1)}{n-1}-(n+1)-(n-k)(n+1)-(n-2)(k-1)\right)\\
+ & \frac{1}{n^2}\left(n^2-1+2\frac{(n-k)(k-1)}{n-1}+\frac{(n-k)^2(n+1)}{n-1}-2(n-k)(n+1)-(k-1)^2\right)\\
= & 2(n-2)\left(s_d^2+2\frac{s_d}{n}\right)+\frac{2k-2}{n^2}.
\end{aligned}
$$

(b) Similarly, we get

$$
\begin{aligned}
\mathbb{V}ar(V_{GMV}|\mathbf{x}) = & V_{GMV;c}^2\left(1+\frac{2}{n+d_0-k-3}+\frac{3}{(n+d_0-k-3)(n+d_0-k-5)}\right)\\
\times & \frac{1}{(n+d_0-k-2)(n+d_0-k-4)} - V_{GMV;c}^2\frac{1}{(n+d_0-k-3)^2}\\
= & 2V_{GMV;c}^2\frac{1}{(n+d_0-k-3)^2(n+d_0-k-5)},
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}ar(R_{GMV}|\mathbf{x}) = & \left(s_c\frac{1}{n+d_0-k-1}+\frac{n+d_0-k-k}{(n+r_0)(n+d_0-k-1)}\right)\frac{V_{GMV;c}}{n+d_0-k-3}\\
+ & V_{GMV;c}\frac{1}{n+d_0-k-3}\left(\frac{n+d_0-k-2}{n+d_0-k-1}s_c+\frac{k-1}{(n+r_0)(n+d_0-k-1)}\right)\\
= & V_{GMV;c}\left(s_c+\frac{1}{n+r_0}\right)\frac{1}{n+d_0-k-3},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{V}ar(s|\mathbf{x}) &= ((n+d_0-k)^2-1)\Bigg(\left(s_c+\frac{1}{n+r_0}\right)^2\\
&+ s_c^2\left(2\frac{n+d_0-k-2}{(n+d_0-k+1)(n+d_0-k-1)^2}+\frac{1}{(n+d_0-k-1)^2}\right)\\
&+ 6\frac{s_c}{n+r_0}\left(-2\frac{n+d_0-k-k}{(n+d_0-k+1)(n+d_0-k-1)^2}+\frac{n+d_0-k-k}{(n+d_0-k-1)^2}\right)\\
&+ \frac{1}{(n+r_0)^2}\left(2\frac{(n+d_0-k-k)(k-1)}{(n+d_0-k+1)(n+d_0-k-1)^2}+\frac{(n+d_0-k-k)^2}{(n+d_0-k-1)^2}\right)\\
&- 2\left(s_c+\frac{1}{n+r_0}\right)\left(s_c\frac{1}{n+d_0-k-1}+\frac{1}{n+r_0}\frac{n+d_0-k-k}{n+d_0-k-1}\right)\Bigg)\\
&- \left((n+d_0-k-2)s_c+\frac{k-1}{n+r_0}\right)^2\\
&= 2(n+d_0-k-2)\left(s_c^2+2\frac{s_c}{n+r_0}\right)+\frac{2k-2}{(n+r_0)^2}.
\end{aligned}
$$

□

Proof of Theorem 16. We present the proof only for the diffuse prior and note that the proof for the conjugate prior is the same. Let $\mathbf{l}=(l_1,l_2,l_3)^\top$ be an arbitrary vector of constant and consider for fixed n

$$
\begin{aligned}
& l_1\sqrt{n}\left(V_{GMV}-\frac{V_{GMV;d}}{n}\right)+l_2\sqrt{n}(R_{GMV}-R_{GMV;d})+l_3\sqrt{n}(s-ns_d)\\
=\ & l_1\left(\frac{V_{GMV;d}}{n}\sqrt{n}\left(\frac{n}{\tau}-1\right)+\frac{V_{GMV;d}}{n}\frac{\psi_1^2/\sqrt{n}}{\phi/n+\varphi/n+\psi_2^2/n}\frac{n}{\tau}\right)\\
+\ & l_2\Bigg(-\sqrt{\frac{V_{GMV;d}}{n}}\frac{\sqrt{ns_d}\psi_2/\sqrt{n}-\sqrt{\phi/n}}{\phi/n+\varphi/n+\psi_2^2/n}\psi_1\\
+\ & \sqrt{\frac{V_{GMV;d}}{n}}\sqrt{1+\frac{\psi_1^2/n}{\phi/n+\varphi/n+\psi_2^2/n}}\frac{\eta}{\sqrt{\tau/n}}\sqrt{ns_d+1-\frac{(\sqrt{ns_d}\psi_2/\sqrt{n}-\sqrt{\phi/n})^2}{\phi/n+\varphi/n+\psi_2^2/n}}\Bigg)\\
+\ & l_3\left(ns_d\sqrt{n}\left(\frac{\xi}{n}-1\right)+\frac{\xi}{n}\left(\frac{\varphi/\sqrt{n}+(1-ns_d)\psi_2^2/\sqrt{n}+2\sqrt{ns_d}\sqrt{\phi/n}\psi_2}{\phi/n+\varphi/n+\psi_2^2/n}\right)\right)
\end{aligned}
$$

The application of the Slutsky theorem (c.f., DasGupta (2008, Theorem 1.5)) and the mutual independence of τ, ψ_1, ψ_2, η, and ξ together with $\sqrt{n}(\xi/n-1)\xrightarrow{d}\mathcal{N}(0,2)$ and $\sqrt{n}(\tau/n-1)\xrightarrow{d}$

$\mathcal{N}(0,2)$ as $n \to \infty$ leads to

$$l_1\sqrt{n}\left(V_{GMV} - \frac{V_{GMV;d}}{n}\right) + l_2\sqrt{n}(R_{GMV} - R_{GMV;d}) + l_3\sqrt{n}(s - ns_d) \xrightarrow{d} \mathcal{N}(0, \sigma_l^2)$$

with

$$\sigma_l^2 \quad = \quad l_1^2 2\breve{V}_{GMV}^2 + l_2^2 \breve{V}_{GMV}\left(1 + \breve{s}\right) + l_3^2\left(2\breve{s}^2 + 4\breve{s}\right) \, .$$

Finally, in using that $\mathbf{l}$ was an arbitrary 3-dimensional vector we get the statement of the theorem. $\square$

Chapter 5

Bayesian Mean-Variance Analysis: Optimal Portfolio Selection under Parameter Uncertainty

The fundamental goal of portfolio theory is to allocate optimally the investments between different assets. The mean-variance optimization is a quantitative tool which allows to make this allocation by considering the trade-off between the risk of portfolio and its return. The basic concepts of modern portfolio theory are developed by Markowitz (1952) who introduced a mean-variance portfolio optimization procedure in which investors incorporate their preferences towards the risk and the expected return to seek the best allocation of wealth. This is attained by selecting the portfolios that maximize the expected portfolio return subject to achieving a prespecified level of risk or, equivalently, minimize the variance subject to achieving a prespecified level of expected return. The mean-variance analysis of Markowitz is an important tool for both practitioners and researchers in financial sector today (see, e.g. Hautsch et al. (2015), Callot et al. (2017)).

The classical problems and pitfalls of the mean-variance analysis are mainly related to extreme weights that often occur when the sample efficient portfolio is constructed. This point was discussed in detail by Merton (1980) who presented an estimator of the instantaneous expected return on the market in a log-normal diffusion price model and showed its slow convergence. Moreover, it was proved that the estimates of the variances and of the covariances of the asset returns are more accurate than the estimates of the means. Best and Grauer (1991) argued that optimal portfolios are very sensitive to the level of expected returns. Therefore, improving the technique of mean estimation has become a key issue of the portfolio optimization problem recently. The same challenge is also present when the covariance matrix need to be estimated. To this end, Broadie (1993) showed that the estimated efficient frontier, a set of all mean-variance optimal portfolios overestimates the expected returns of portfolios for different levels of estimation errors. A similar conclusion has also been drawn in more recent studies by Basak et al. (2005); Siegel and Woodgate (2007); Bodnar and Bodnar (2010).

An alternative approach to deal with the parameter uncertainty in portfolio analysis is to employ the methods of Bayesian statistics (c.f., Barry (1974), Brown (1976), Klein and Bawa (1976), Frost and Savarino (1986), Aguilar and West (2000), Rachev et al. (2008), Avramov and Zhou (2010), Sekerke (2015), Bodnar et al. (2017b)). It is remarkable that the Bayesian approach is potentially more attractive since i) it uses prior information about quantities of interest; ii) it facilitates the use of fast, intuitive, and easily implementable numerical algorithms in order to simulate complex economic quantities; iii) it accounts for estimation risk and model uncertainty in the portfolio choice problem. First applications of Bayesian statistics to portfolio analysis during the 1970s were completely based on noninformative or data-based priors. Bawa et al. (1979) provided an excellent early survey on such applications. The Bayesian approaches which are based on the diffusion prior are usually comparable with the classical methods for the portfolio selection. However, if some of the risky assets have longer histories than other, then the Bayesian approaches under the diffuse prior lead to different results (see Stambaugh (1997)). Jorion (1986) introduced the hyperparameter prior approach in the spirit of the Bayes-Stein shrinkage prior, whereas Black and Litterman (1992) defended an informal Bayesian analysis with economic arguments and equilibrium relations. They derived the Black-Litterman model which leads to more stable and more diversified portfolios than simple mean-variance optimization. Unfortunately, the application of this model requires a broad variety of data, some of which may be hard to find. Recent studies by Pástor (2000) and Pástor and Stambaugh (2000) centered prior beliefs around values implied by asset pricing theories. In particular, Pástor and Stambaugh (2000) investigated the portfolio choices of mean-variance-optimizing investors who use sample evidence to update prior beliefs centered on either risk-based or characteristic-based pricing models. Tu and Zhou (2010) argued that the investment objective provides a useful prior for portfolio selection and proposed an optimal combination of the naive equally weighted portfolio rule with one of the four sophisticated strategies – the Markowitz rule, the Jorion (1986) rule, the MacKinlay and Pástor (2000) rule, and the Kan and Zhou (2007) rule – as a way to improve the performance.

We contribute to the existent literature of optimal portfolio selection by formulating the optimization problem in terms of the posterior predictive distribution and solving it. Using the available information about the development of asset returns which is present in their historical observations, the aim is to construct an optimal portfolio by taking into account investor's preferences. The conventional approach consist of two steps: (i) first, the optimization problem is solved with the solution depending on the unknown parameters of the asset return distribution; (ii) second, the optimal portfolio weights, which are the solutions of optimization problem, are estimated by applying the historical observations of the asset returns. It is important to note that following this approach, the obtained solution is sub-optimal only and it can deviate considerably from the optimal (population) portfolio obtained in the first stage.

In this chapter, we propose a new approach, where the solution of the investor's optimization problem is obtained by employing the posterior predictive distribution which takes parameter uncertainty into account before the optimal portfolio choice problem is solved. As a result, its solution is present in terms of historical data and is independent of unknown parameters of the asset return distribution. Consequently, it can be directly applied in practice and, in contrast to the conventional approach, it is optimal.

The rest of the chapter is organized as follows. Main theoretical results are given in section 5.1. Here, we characterize the posterior predictive distribution of the asset return by developing a very helpful stochastic representation (Theorem 17). This stochastic representation provides not only a way how future realization of portfolio returns could be simulated, but also it is used to derive the first two moments needed in the considered optimization problem. Section 5.1.2 deals with constructing optimal portfolios by maximizing the posterior mean-variance utility function, while the expression of the Bayesian efficient frontier is derived in section 5.1.3. The theoretical results are implement in an empirical study of sSection 5.2, while section 5.3 provides a conclusion. The technical derivations are moved to section 5.4.

5.1 Mean-variance analysis under parameter uncertainty

5.1.1 Posterior predictive distribution

Let $\mathbf{X}_t$ denotes the k-dimensional vector of returns on asset at time t. Assume that a sample of size n of asset returns $\mathbf{x}_{t-n}, ..., \mathbf{x}_{t-1}$, realizations of $\mathbf{X}_{t-n}, ..., \mathbf{X}_{t-1}$, is available which provides the information set $\mathcal{F}_t$ and let $\mathbf{x}_{(t-1)} = (\mathbf{x}_{t-n}, ..., \mathbf{x}_{t-1})$ be the observation matrix at time $t-1$. Consequently, an investor makes a decision by optimising preferences using information $\mathcal{F}_t$.

Before the decision problem is formulated in section 5.1.2, we first derive the predictive posterior distribution $p(\mathbf{X}_t|\mathbf{x}_{(t-1)})$ of $\mathbf{X}_t$ given the previous observation of asset returns summarized in $\mathbf{x}_{(t-1)}$. The derivation of $p(\mathbf{X}_t|\mathbf{x}_{(t-1)})$ is based on the methods of Bayesian statistics which provide well-established techniques for providing inferences of future realizations of asset returns given information $\mathcal{F}_t$.

In the following we assume that the asset returns $\mathbf{X}_1, \mathbf{X}_2, ...$ are infinitely exchangeable and multivariate centered spherically symmetric (see, Bernardo and Smith (2000, Section 4.4) for the definition and properties). This assumption is very general and it implies that neither the unconditional distribution of the asset returns is normal nor that they are independently distributed. Moreover, the unconditional distribution of the asset returns appears to be heavy-tailed which is usually observed for financial data (see, e.g., Bradley and Taqqu (2003)).

Parameterizing the density function of $\mathbf{X}_{(t-1)} = (\mathbf{X}_{t-n}, ..., \mathbf{X}_{t-1})$ by the parameter $\boldsymbol{\theta}$, the posterior distribution of $\boldsymbol{\theta}$ is obtained by applying the Bayes theorem and it is given by

$$\pi(\boldsymbol{\theta}|\mathbf{x}_{(t-1)}) \propto f(\mathbf{x}_{(t-1)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{5.1}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior and $f(\mathbf{x}_{(t-1)}|\boldsymbol{\theta})$ is the likelihood function of $\mathbf{X}_{(t-1)}$. The posterior distribution $\boldsymbol{\theta}$ is then used to derive the posterior predictive distribution of the portfolio return at time t expressed as

$$X_{p,t} = \boldsymbol{w}^\top \mathbf{X}_t, \tag{5.2}$$

where $\boldsymbol{w} = (w_1, ..., w_p)^\top$ is the k-dimensional vector of portfolio weights.

The posterior distribution (5.1) is employed in the derivation of the posterior predictive distribution as follows:

$$f(x_{p,t}|\mathbf{x}_{(t-1)}) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} f(x_{p,t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x}_{(t-1)})\mathrm{d}\boldsymbol{\theta}\,. \tag{5.3}$$

Due to the integration present in the definition of the posterior predictive distribution, it is possible to obtaine the analytical expression of $f(x_{p,t}|\mathbf{x}_{(t-1)})$ only in very rare cases. Moreover, the integration in (5.3) could also be high-dimensional, which makes the application of numerical methods very time consuming and also questions the quality of their numerical approximation. In Theorem 17, we derive a stochastic representation for the posterior predictive distribution $f(x_{p,t}|\mathbf{x}_{(t-1)})$ which can be very easily used to draw sample from this distribution as well as to compute its expected value and variance analytically. Finally, it has to be noted that the application of the stochastic representation describing the distribution of random quantities has been used both in the frequentist statistics (see, e.g., Givens and Hoeting (2012), Gupta et al. (2013)) and the Bayesian statistics (c.f., Bodnar et al. (2017b)).

Theorem 17. *Let $\mathbf{X}_1, \mathbf{X}_2, ...$ are infinitely exchangeable and multivariate centered spherically symmetric. Let $\pi(\boldsymbol{\theta}) = |\mathbf{F}|^{1/2}$ be Jeffreys' prior where $|\mathbf{A}|$ denotes the determinant of a square matrix $\mathbf{A}$ and $\mathbf{F} = -\mathbb{E}\left(\frac{\partial^2 \log(f(\mathbf{x}_{(t-1)}|\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right)$ is the Fisher information matrix. Assume $n > k$. Then the stochastic representation of the random variable $\widehat{X}_{p,t}$ whose density is the posterior predictive distribution (5.3) is given by*

$$\widehat{X}_{p,t} \;\overset{d}{=}\; \boldsymbol{w}^\top \overline{\mathbf{x}}_{t-1} + \sqrt{\boldsymbol{w}^\top \mathbf{S}_{t-1}\boldsymbol{w}}\left(\frac{t_1}{\sqrt{n(n-k)}} + \sqrt{1+\frac{t_1^2}{n-k}}\frac{t_2}{\sqrt{n-k+1}}\right),$$

where

$$\overline{\mathbf{x}}_{t-1} = \frac{1}{n}\sum_{i=t-n}^{t-1} \mathbf{x}_i \quad \textit{and} \quad \mathbf{S}_{t-1} = \sum_{i=t-n}^{t-1} (\mathbf{x}_i - \overline{\mathbf{x}}_t)(\mathbf{x}_i - \overline{\mathbf{x}}_t)^\top. \tag{5.4}$$

and t_1, t_2 *are independent with* $t_1 \sim t_{n-k}$ *and* $t_2 \sim t_{n-k+1}$. *The symbol "*$\stackrel{d}{=}$*" denotes the equality in distribution.*

The result of Theorem 17 provide an easy way how a random sample from the posterior distribution of $f(\mathbf{x}_t|\mathbf{x}_{(t-1)})$ can be simulated:

(i) generate $t_1^{(b)} \sim t_{n-k}$ and $t_2^{(b)} \sim t_{n-k+1}$;

(ii) compute

$$\widehat{X}_{p,t}^{(b)} = \boldsymbol{w}^\top \overline{\mathbf{x}}_t + \sqrt{\boldsymbol{w}^\top \mathbf{S}_t \boldsymbol{w}} \left(\frac{t_1^{(b)}}{\sqrt{n(n-k)}} + \sqrt{1 + \frac{(t_1^{(b)})^2}{n-k}} \frac{t_2^{(b)}}{\sqrt{n-k+1}} \right)$$

(iii) Repeat steps (i) and (ii) for $b = 1, ..., B$ resulting in independent sample $\widehat{X}_{p,t}^{(1)}, ..., \widehat{X}_{p,t}^{(B)}$ from the posterior predictive distribution (5.3).

The generated sample $\widehat{X}_{p,t}^{(1)}, ..., \widehat{X}_{p,t}^{(B)}$ is the used to calculate important characteristics of the distribution $f(\mathbf{x}_t|\mathbf{x}_{(t-1)})$, like the mean, the variance, the credible interval, etc. To this end, we note that the condition $n > k$ ensures that $\mathbf{S}_t$ is positive definite and, hence, it is invertible.

Another important application of Theorem 17 provides us with the analytical expression of the expected value and the variance of the posterior predictive distribution $f(\mathbf{x}_t|\mathbf{x}_{(t-1)})$. These findings are formulated in Corollary 2

Corollary 2. *Under the conditions of Theorem 17, let* $n - k > 2$. *Then:*

$$\mathbb{E}(\mathbf{X}_t|\mathbf{x}_{(t-1)}) = \boldsymbol{w}^\top \overline{\mathbf{x}}_{t-1} \tag{5.5}$$

and

$$\mathbb{V}ar(\mathbf{X}_t|\mathbf{x}_{(t-1)}) = c_{k,n} \boldsymbol{w}^\top \mathbf{S}_{t-1} \boldsymbol{w} \quad \textit{with} \quad c_{k,n} = \frac{1}{n-k-1} + \frac{2n-k-1}{n(n-k-1)(n-k-2)} \tag{5.6}$$

The proof of Corollary 2 is given in section 5.4. Its results are used in the next section, where the expressions of optimal portfolio weights are given.

5.1.2 Mean-variance optimal portfolios

The mean-variance investor constructs an optimal portfolio at time $t-1$ for the next period by maximizing the mean-variance utility function given by

$$U(\boldsymbol{w}) = \mathbb{E}(\mathbf{X}_t|\mathbf{x}_{(t-1)}) - \frac{\gamma}{2}\mathbb{V}ar(\mathbf{X}_t|\mathbf{x}_{(t-1)}) = \boldsymbol{w}^\top\overline{\mathbf{x}}_{t-1} - \frac{c_{k,n}\gamma}{2}\boldsymbol{w}^\top\mathbf{S}_{t-1}\boldsymbol{w} \tag{5.7}$$

under the constraint that the whole wealth is invested into the selected assets, i.e., $\boldsymbol{w}^\top\mathbf{1} = 1$ where $\mathbf{1}$ denotes the k-dimensional vector of ones. The quantity $\gamma > 0$ stands for the coefficient of the investor's risk aversion and describes the investor's attitude towards risk.

In contrast to the conventional approach that involves the unknown parameters of the asset return distribution in its formulation, the optimization problem in (5.7) already incorporates the parameter uncertainty by using the available information summarized in the data matrix $\mathbf{x}_{(t-1)}$. As a result, the output of solving (5.7) is the formula for optimal portfolio weights that could be directly applied in practice, while the estimation of optimal portfolio weights is required in the conventional methods that leads to the suboptimality of the resulting portfolio.

The optimization problem in (5.7) is similar to the optimization problem in the conventional approach (see Ingersoll (1987); Okhrin and Schmid (2006)) with the exception that the risk aversion coefficient is multiplied by the constant $c_{k,n}$. As a results, the solution of (5.7) is given by

$$\boldsymbol{w}_{MV,\gamma} = \frac{\mathbf{S}_{t-1}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \gamma^{-1}c_{k,n}^{-1}\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1} \quad \text{with} \quad \mathbf{Q}_{t-1} = \mathbf{S}_{t-1}^{-1} - \frac{\mathbf{S}_{t-1}^{-1}\mathbf{1}\mathbf{1}'\mathbf{S}_{t-1}^{-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} \tag{5.8}$$

together with the expected return and the variance expressed as

$$R_{MV,\gamma} = \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \gamma^{-1}c_{k,n}^{-1}\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1} \tag{5.9}$$

and

$$V_{MV,\gamma} = \frac{c_{k,n}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \gamma^{-2}c_{k,n}^{-1}\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}, \tag{5.10}$$

respectively, where we use that $\mathbf{Q}_{t-1}\mathbf{1} = \mathbf{0}$ and $\mathbf{Q}_{t-1}\mathbf{S}_{t-1}\mathbf{Q}_{t-1} = \mathbf{Q}_{t-1}$ in (5.10).

Additionally to the formulae of the optimal portfolio weights, the expected return and the variance of the mean-variance optimal portfolios presented in (5.8)-(5.10), the Bayesian approach allows to characterize the posterior predictive distribution of the constructed optimal portfolio. This is achieved by applying the results of Theorem 17 where the weights of an arbitrary portfolio are replaced by the optimal portfolio weights given in (5.8). Then, the posterior predictive distribution of the optimal portfolio return is obtained via simulations as described after Theorem 17 by replacing $\boldsymbol{w}$ with $\boldsymbol{w}_{MV,\gamma}$ as in (5.8). This is a very important result which allows the whole characterization of the stochastic behaviour of optimal portfolio return and is a great advantage with respect to the conventional approach where the point estimator is only present.

We conclude this section by noting that the original Markowitz problem (see Markowitz (1952, 1959)) is solved in the same way. In the mean variance analysis of Markowitz, the optimization problem is given by: (i) minimizing the portfolio variance for a given level of the expected return R_0 or (ii) maximizing the expected return for the given level of the variance V_0. In the first case the optimal portfolio weights are given by

$$\boldsymbol{w}_{MV,R_0} = \frac{\mathbf{S}_{t-1}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \left(R_0 - \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}}\right)\frac{\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}{\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}} \tag{5.11}$$

with

$$V_{MV,R_0} = \frac{c_{k,n}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + c_{k,n}\left(R_0 - \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}}\right)^2\frac{1}{\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}, \tag{5.12}$$

while the solution of the second optimization problem is

$$\boldsymbol{w}_{MV,V_0} = \frac{\mathbf{S}_{t-1}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \sqrt{c_{k,n}^{-1}V_0 - \frac{1}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}}}\frac{\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}{\sqrt{\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}} \tag{5.13}$$

with

$$R_{MV,V_0} = \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \sqrt{c_{k,n}^{-1}V_0 - \frac{1}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}}}\sqrt{\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}. \tag{5.14}$$

5.1.3 Bayesian efficient frontier

Equations (5.9) and (5.10) determine the set of all optimal portfolios obtained as solutions of (5.7) for $\gamma > 0$. Solving these two equation with respect to γ leads to a set in the mean-variance space where all mean-variance optimal portfolios lie. We call this set the Bayesian efficient frontier which is given by

$$(R - R_{GMV})^2 = \frac{\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}{c_{k,n}}\left(V - V_{GMV}\right), \tag{5.15}$$

where

$$R_{GMV} = \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} \quad \text{and} \quad V_{GMV} = \frac{c_{k,n}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} \tag{5.16}$$

are the expected return of the global minimum variance portfolio, i.e., the mean-variance optimal portfolio with the smallest variance, with the weights expressed as

$$\boldsymbol{w}_{GMV} = \frac{\mathbf{S}_{t-1}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}}. \tag{5.17}$$

The quantity $s = \overline{\mathbf{x}}_{t-1}^{\top}\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}/c_{k,n}$ is the slope parameter of the efficient frontier which is equal to the amount of the excess squared return with respect to the return of the global minimum variance portfolio when the variance is increased by one. Finally, we note that the Bayesian efficient frontier is a parabola in the mean-variance space which is the same finding as obtained by the conventional approach (see Merton (1972)).

5.2 Empirical illustration

5.2.1 Data

For an empirical illustration, we use weekly returns from a collection of assets of the S&P500, allowing for portfolios ranging from 5 to 40 assets. The parameters are estimated with sample sizes of $n \in \{52, 78, 104, 130\}$, corresponding to one year up to two and a half years of weekly data. All the data end on the 8th of October 2017. For $n = 52$, this corresponds to almost the whole presidency of Donald Trump, which was, regarding the S&P500, a period of almost stable growth from 2200 to 2600 points. But besides of two slight drops in August 2015 and the early weeks of 2016, this holds for the other periods - despite of Trump's presidency. The constructed portfolios consist of $k \in \{5, 10, 25, 40\}$ assets. This allows us to analyze the behaviour of the proposed model not only in terms of economic risk but also regarding statistical estimation uncertainty.

5.2.2 Conventional approach

Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the mean vector and the covariance matrix of the asset returns. Then the traditional approach to construct an optimal portfolio consists of two steps (see, e.g., Ingersoll (1987); Okhrin and Schmid (2006)):

(1) The optimization problem

$$\boldsymbol{w}^{\top}\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^{\top}\boldsymbol{\Sigma}\boldsymbol{w} \longrightarrow \max \quad \text{subjct to} \quad \boldsymbol{w}^{\top}\mathbf{1} = 1 \tag{5.18}$$

is solved resulting in the expression of optimal portfolio weights presented in terms of the population (unknown) parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$\boldsymbol{w}_{P,\gamma} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \gamma^{-1}\mathbf{R}\boldsymbol{\mu} \quad \text{with} \quad \mathbf{R} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} \tag{5.19}$$

with the expected return and the variance expressed as

$$R_{P,\gamma} = \frac{\mathbf{1}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \gamma^{-1}\boldsymbol{\mu}^\top\mathbf{R}\boldsymbol{\mu} \quad \text{and} \quad V_{P,\gamma} = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} + \gamma^{-2}\boldsymbol{\mu}^\top\mathbf{R}\boldsymbol{\mu}, \tag{5.20}$$

(2) The unknown population quantities are replaced by their sample counterparts, i.e. by the sample mean vector and the sample covariance matrix given by

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_{t-1} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = d_n\mathbf{S}_{t-1} \quad \text{with} \quad d_n = \frac{1}{n-1}$$

Then the sample optimal portfolio weights are obtained by

$$\boldsymbol{w}_{S,\gamma} = \frac{\mathbf{S}_{t-1}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \gamma^{-1}d_n^{-1}\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1} \tag{5.21}$$

with the sample estimators for the expected return and for the variance given by

$$R_{S,\gamma} = \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \gamma^{-1}d_n^{-1}\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1} \quad \text{and} \quad V_{S,\gamma} = \frac{d_n}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} + \gamma^{-2}d_n^{-1}\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}\,. \tag{5.22}$$

In the similar way, the sample efficient frontier is constructed by (see Bodnar and Schmid (2008b, 2009); Kan and Smith (2008))

$$(R - R_{GMV,S})^2 = \frac{\overline{\mathbf{x}}_{t-1}^\top\mathbf{Q}_{t-1}\overline{\mathbf{x}}_{t-1}}{d_n}\left(V - V_{GMV,S}\right), \tag{5.23}$$

where

$$R_{GMV,S} = \frac{\mathbf{1}^\top\mathbf{S}_{t-1}^{-1}\overline{\mathbf{x}}_{t-1}}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} \quad \text{and} \quad V_{GMV,S} = \frac{d_n}{\mathbf{1}'\mathbf{S}_{t-1}^{-1}\mathbf{1}} \tag{5.24}$$

which is an estimator of the population efficient frontier.

It is remarkable that the expression of the sample optimal portfolio weights has the same structure as the weights of the optimal portfolios obtained following the Bayesian approach. The only difference is that $c_{k,n}$ in (5.8) is replaced by d_n in (5.21). Similar results are also obtained

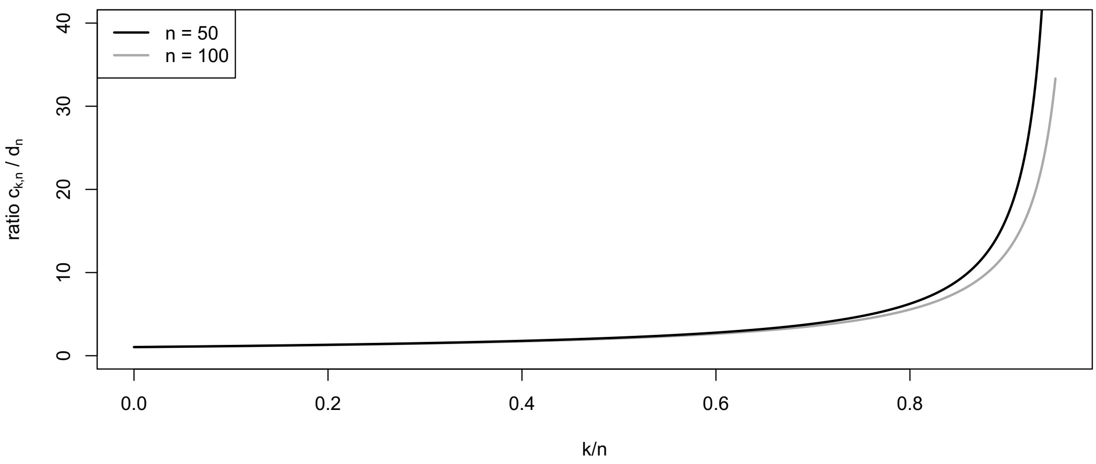

Figure 5.1: The ratio $c_{k,n}/d_n$ plotted as a function of k/n for $k/n \in [0, 0.95)$ and $n \in \{50, 100\}$.

in the case of the efficient frontier which is fully determined by three parameters: the mean and the variance of the global minimum variance portfolio and the slope parameter. While the formulae in the case of the mean of the global minimum variance portfolio coincide, this is not longer true for the variance of the global minimum variance portfolio and the slope coefficient. The Bayesian approach leads to a larger value of the variance and to a smaller value of the slope parameter. The difference between the corresponding expressions obtained by the sample estimation or derived from the Bayeian posterior distribution as in Section 2 can be considerable when the portfolio dimension is comparable to the sample size as shown in Figure 5.1, where we plot the ratio $c_{k,n}/d_n$ as a function of k/n for $n \in \{50, 100\}$. We observe that when the number of assets k gets closer to the sample size, even for a moderate ratio of $k/n = 0.6$, the Bayesian estimator and the sample estimator deviate. If the number of assets corresponds almost to the sample size, the estimators deviate considerably. Since it is sometimes necessary to restrict an estimation to a smaller sample size, e.g. after a structural break in the data, the difference in the estimators has to be considered.

It is a well-known fact that the sample efficient frontier is overoptimistic and overestimates the location of the population efficient frontier in the mean-variance (c.f., Basak et al. (2005); Siegel and Woodgate (2007); Bodnar and Bodnar (2010)). In contrast, the Bayesian approach provides an improved procedure which shrinks the sample efficient frontier by increasing the estimated variance of the global minimum portfolio and reducing the slope parameter. We illustrate this point in Section 5.2.3 on real data described in Section 5.2.1.

5.2.3 Comparison study

As mentioned in the previous section, there is a distinct difference between the classical sample estimators and the Bayesian estimators proposed in this paper. With this conclusion and the fact that the sample efficient frontier overestimates the population efficient frontier, we expect the estimations for the return and the variance to be larger in the Bayesian case compared to the sample estimations indicating that the Bayesian approach also takes the estimation risk into account in its construction which in practice automatically leads to smaller values of the risk aversion coefficient in comparison to the conventional case. Figure 5.2 illustrates this presumption: fixing $n = 130$ and considering different portfolio sizes $k \in \{5, 10, 25, 40\}$ for different risk attitudes $\gamma \in \{10, 25, 50, 100\}$, we find that for the same value of the risk coefficient γ and for the same portfolio size, the Bayesian estimator performs as expected compared to the sample estimator. Furthermore, the difference in the estimators increases if the number of assets gets closer to the sample size, as illustrated in Figure 5.1 or when γ decreases, i.e. for less risk averse investors the impact of parameter uncertainty becomes larger.

Regarding the efficient frontier, Figure 5.3 shows the estimated efficient frontiers for a fixed sample size of $n = 130$ and varying portfolio sizes $k \in \{5, 10, 25, 40\}$ in the Bayesian case as well as the conventional case. The Bayesian efficient frontier lies always below the sample efficient frontier and therefore exhibits less overestimation of the population efficient frontier. Furthermore, Figure 5.3 also illustrates the finding shown in Figure 5.1. The estimators of the efficient frontier deviate stronger when the portfolio size gets closer to the sample size. This fact is also illustrated in Figure 5.4 for fixed $k = 40$ and varying $n \in \{52, 78, 104, 130\}$. The two estimated efficient frontiers coincide more the larger the sample size n is. This is in line with the theoretical implications. Finally, we also observe the increase in the slope parameter of the efficient frontier when the portfolio dimension increases indicating the well-documented positive effect of portfolio diversification.

5.2.4 Posterior interval prediction

In contrast to the conventional procedure, the Bayesian approach provides also the whole posterior predictive distribution of the constructed optimal portfolio return and not only the point estimator of its weights. Using data described in section 5.2.1, we calculated in this section the prediction intervals for the optimal portfolio returns calculated for several values of the risk-aversion coefficient $\gamma \in \{10, 20, ..., 100\}$, for $k \in \{5, 25\}$, and for $n \in \{52, 78, 104, 130\}$ (see Figure 5.5).

The prediction intervals in Figure 5.5 are obtained as follows:

(a) Fix γ and calculate the expected return and the variance of the corresponding mean-variance optimal portfolio as given (5.9) and (5.10);

(b) For chosen γ, compute the weights of the optimal mean-variance portfolio $\boldsymbol{w}_{MV,\gamma}$ using (5.8).

(c) In using $\boldsymbol{w}_{MV,\gamma}$ apply the results of Theorem 17 and the simulation procedure described after the statement of this theorem to get a sample of optimal portfolio returns denoted by $R^{(b)}_{MV,\gamma}$ for $b = 1, ..., B$.

(d) Fix the significance level of the prediction interval α and compute the $\alpha/2$- and $(1-\alpha/2)$-quantiles from the empirical distribution of $R^{(b)}_{MV,\gamma}$, $b = 1, ..., B$

(e) For the computed value of $V_{GM,\gamma}$ in part (a), plot the point prediction $R_{GM,\gamma}$ from (a) together with the prediction interval from (d).

The order of the efficient portfolios given in Figure 5.5 is directly determined by the risk aversion coefficient. The smaller γ, the riskier is the portfolio and lies therefore more right on the efficient frontier. We observe that the optimal efficient portfolios are shifted to the right for growing sample sizes. But the focus lies here on the credible intervals for a confidence level of $\alpha = 0.05$. The first observation is that no credible interval covers negative values, implying positive portfolio returns with probability of 95%. The second observation is that the credible intervals become larger the more risky an efficient portfolio becomes – which is in line with the theory. And the third observation is that these credible intervals for riskier efficient portfolios become larger regardless of the increased sample size. Hence, the decrease in estimation risk resulting from a larger sample is outweighed by the economic risk.

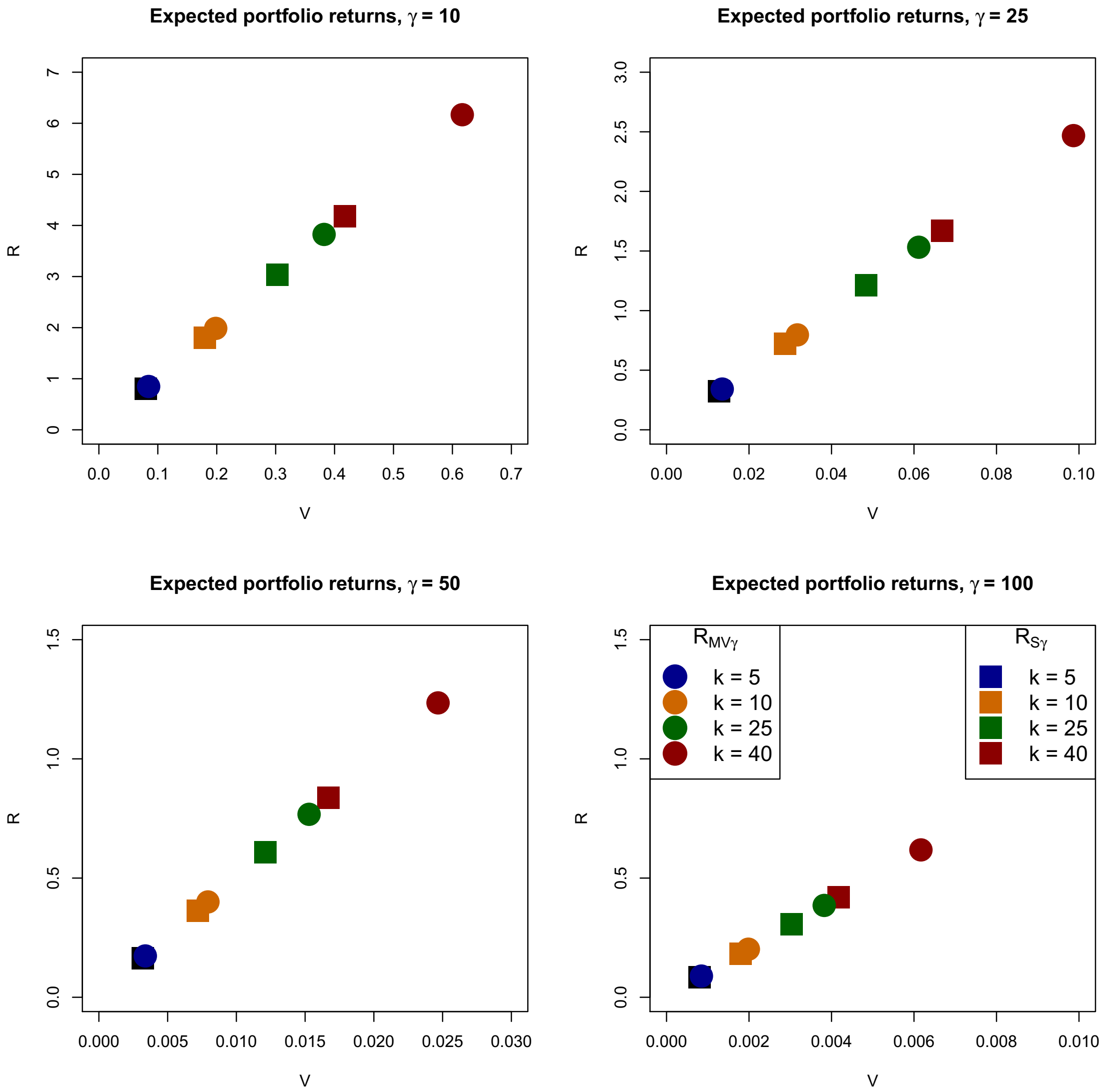

Figure 5.2: Sample optimal portfolios and Bayesian optimal portfolios.
For the risk aversion coefficient of $\gamma \in \{10, 25, 50, 100\}$, for the sample case of $n = 130$ and for the portfolio dimension of $k \in \{5, 10, 25, 40\}$.

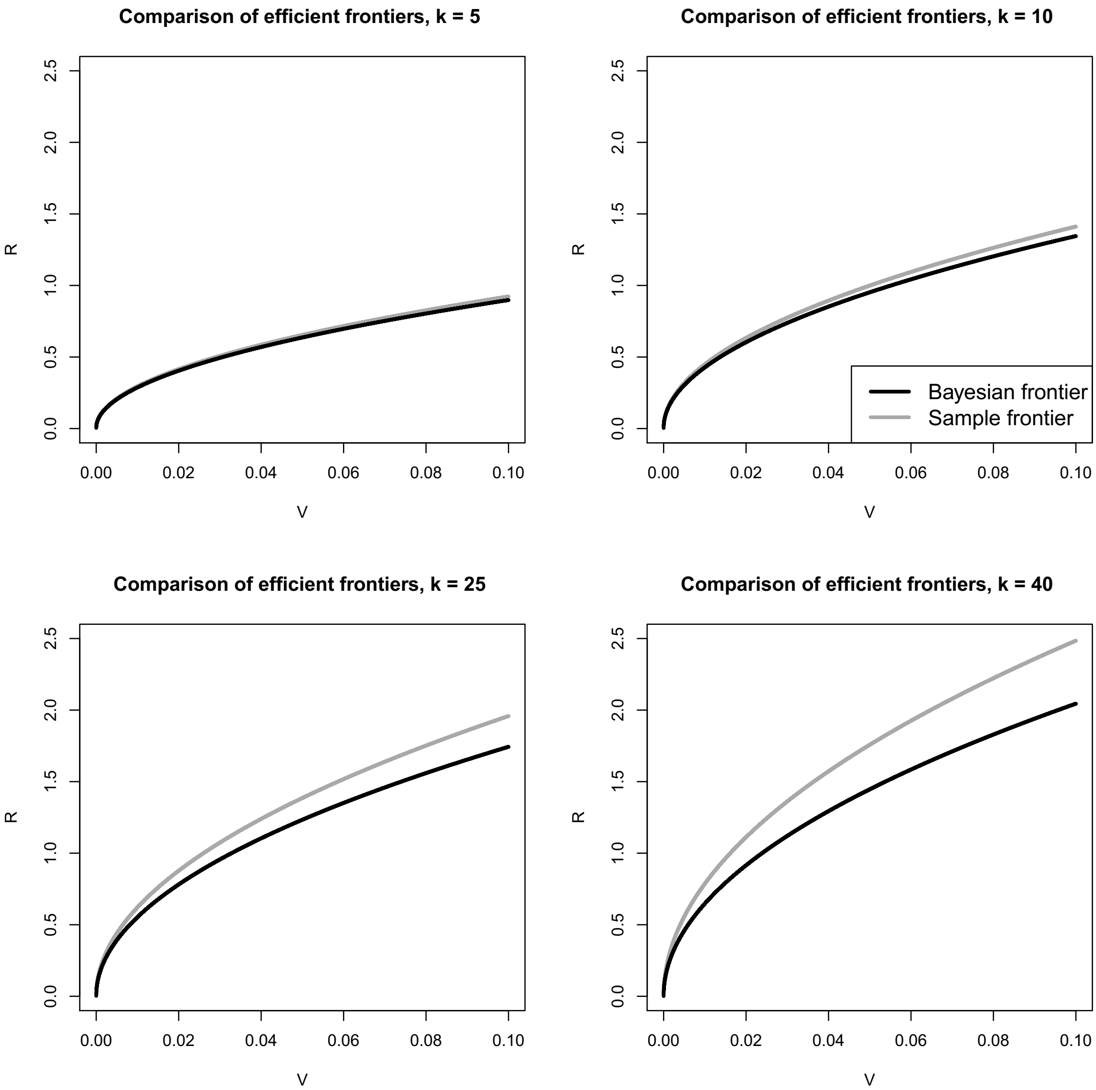

Figure 5.3: The sample efficient frontiers and the Bayesian efficient frontier. $n = 130$ and $k \in \{5, 10, 25, 40\}$.

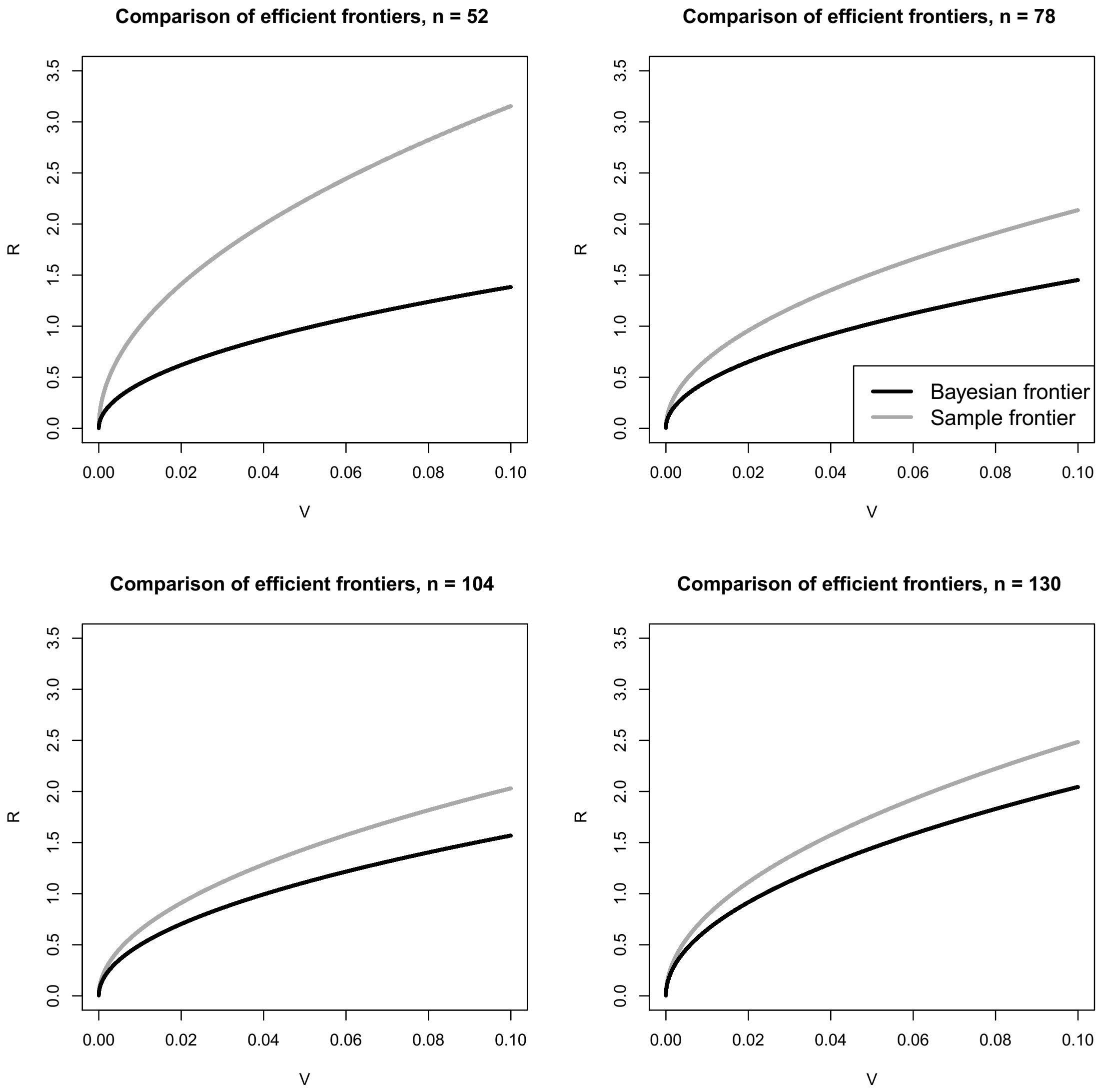

Figure 5.4: The sample efficient frontiers and the Bayesian efficient frontier. $k = 40$ and $n \in \{52, 78, 104, 130\}$.

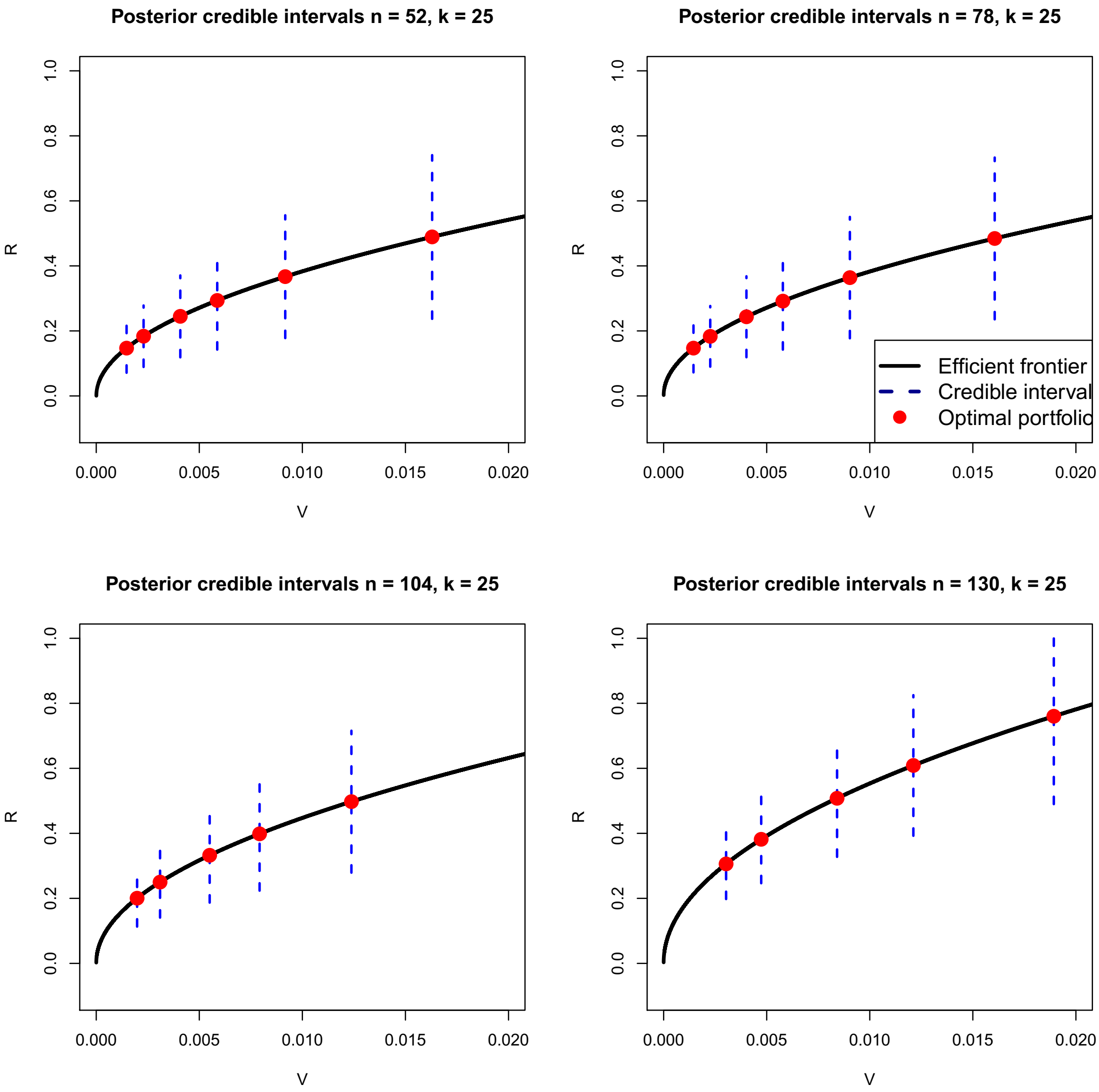

Figure 5.5: Credible intervals for the return of optimal portfolios with varying risk attitudes. The sample sizes are chosen to be $n \in \{52, 78, 104, 130\}$ and the portfolio size is fixed to $k = 25$. The confidence level is set to $\alpha = 0.05$.

5.3 Conclusion

The mean-variance analysis of Markowitz presents a fundamental way of portfolio construction which is very popular in the financial literature today. It provides an investor the portfolio weights which determine the structure of the optimal portfolio. However, the investor faces with a number of difficulties by implementing this procedure in practice. One of the main pitfalls of the mean-variance analysis is that its solution is presented in terms of unobservable quantities, the parameters of the asset returns distribution. As a results, the optimization problem is performed in two steps. After finding the analytical solution, the optimal portfolio is constructed by replacing the unknown parameters with their estimates. Due to the considerable influence of parameter uncertainty on the investment process, this procedure leads only to sub-optimal portfolios.

We deal with the problem from the viewpoint of Bayesian statistics. The optimization problem is formulated in terms of the posterior predictive distribution which does not involve unknown quantities. Consequently, we deal with parameter uncertainty before solving the optimization problems. This approach allows us to find optimal portfolio weights which now depend only on historical observations of the asset returns. The advantages of the approach are shown both theoretically and empirically. In particular, we show that the constructed Bayesian efficient frontier improves the overoptimism which is present in the sample efficient frontier. Another important advantage of the suggested procedure is that it allows us not only to construct an optimal portfolio based on the posterior predictive distribution, but also an intelligent technique in performing an interval forecast of future realizations of optimal portfolio returns which are obtained by employing the derived stochastic representation of the posterior predictive distribution.

5.4 Proofs and Supplementary Material

Proof of Theorem 17: The assumptions of infinitely exchangeability and multivariate centered spherically symmetry implies (see, e.g., Bernardo and Smith (2000, Proposition 4.6)) that the asset returns are independently and identically distributed given the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ with the conditional distribution given by $\mathbf{X}_t|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (k-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$). Under this model with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Jeffreys' prior is given by

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \propto \quad |\boldsymbol{\Sigma}|^{-(k+1)/2}, \tag{5.25}$$

which leads to the posterior expressed as

$$\pi(\boldsymbol{\mu},\boldsymbol{\Sigma}|\mathbf{x}_{(t-1)}) \propto |\boldsymbol{\Sigma}|^{-(n+k+1)/2}\exp\left\{-\frac{n}{2}(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})-\frac{1}{2}\mathrm{tr}[\mathbf{S}_{t-1}\boldsymbol{\Sigma}^{-1}]\right\}, \quad (5.26)$$

where $\overline{\mathbf{x}}_{t-1}$ and $\mathbf{S}_{t-1}$ are given in the statement of the theorem.

From (5.26) we obtain that the posterior distribution of $\boldsymbol{\Sigma}$ is the inverse Wishart distribution (see Gupta and Nagar (2000) for the definition and properties) given by

$$\boldsymbol{\Sigma}|\boldsymbol{\mu},\mathbf{x}_{t-1} \sim \mathcal{IW}_k(n+k+1,\widetilde{\mathbf{S}}_{t-1}(\boldsymbol{\mu})) \text{ with } \widetilde{\mathbf{S}}_{t-1}(\boldsymbol{\mu}) = \mathbf{S}_{t-1}+n(\boldsymbol{\mu}-\overline{\mathbf{x}}_{t-1})(\boldsymbol{\mu}-\overline{\mathbf{x}}_{t-1})^\top. \quad (5.27)$$

Furthermore, integrating out $\boldsymbol{\Sigma}$ we get the marginal posterior for $\boldsymbol{\mu}$ expressed as

$$\begin{aligned}\pi(\boldsymbol{\mu}|\mathbf{x}_{(t-1)}) &\propto \int_{\boldsymbol{\Sigma}>0}|\boldsymbol{\Sigma}|^{-(n+k+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[(n(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})^\top+\mathbf{S}_{t-1})\boldsymbol{\Sigma}^{-1}\right]\right\}d\boldsymbol{\Sigma}\\ &\propto |n(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})^\top+\mathbf{S}_{t-1}|^{-\frac{n}{2}},\end{aligned}$$

where the last equality follows by observing that the function under the integral is the density function of the inverse Wishart distribution with $n+k+1$ degrees of freedom and parameter matrix $n(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})^\top+\mathbf{S}_{t-1}$. The application of Silvester's determinant theorem leads to

$$\pi(\boldsymbol{\mu}|\mathbf{x}_{(t-1)}) \propto \left(1+n(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})^\top\mathbf{S}_{t-1}^{-1}(\overline{\mathbf{x}}_{t-1}-\boldsymbol{\mu})\right)^{-\frac{n}{2}}, \quad (5.28)$$

which proves that $\boldsymbol{\mu}|\mathbf{x}_{t-1} \sim t_k\left(n-k,\overline{\mathbf{x}}_{t-1},\frac{1}{n(n-k)}\mathbf{S}_{t-1}\right)$ (k-dimensional multivariate t-distribution with $n-k$ degrees of freedom, location vector $\overline{\mathbf{x}}_{t-1}$, and scale matrix $\frac{1}{n(n-k)}\mathbf{S}_{t-1}$).

Because $\mathbf{X}_{t-n},...,\mathbf{X}_t$ are independent given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as well as conditionally normally distributed, we get that the conditional distribution $X_{p,t}|\boldsymbol{\mu},\boldsymbol{\Sigma}$ coincides with $X_{p,t}|\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{x}_{(t-1)}$ given by

$$X_{p,t}|\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{x}_{(t-1)} \sim \mathcal{N}(\boldsymbol{w}^\top\boldsymbol{\mu},\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}),$$

where the last equality proves that $X_{p,t}$ depends on $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\mathbf{x}_{(t-1)}$ only over $\boldsymbol{w}^\top\boldsymbol{\mu}$ and $\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}$.

The application of Theorem 3.2.13 in Muirhead (1982) leads to

$$\frac{\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}}{\boldsymbol{w}^\top\widetilde{\mathbf{S}}_{t-1}(\boldsymbol{\mu})\boldsymbol{w}} \stackrel{d}{=} \frac{1}{\xi}, \quad (5.29)$$

where $\xi \sim \chi^2_{n-k+1}$ and is independent of $\boldsymbol{\mu}$ and $\mathbf{X}_{(t-1)}$. Then the stochastic representation of

$X_{p,t}$ is given by

$$X_{p,t} \stackrel{d}{=} \boldsymbol{w}^\top \boldsymbol{\mu} + \frac{\sqrt{\boldsymbol{w}^\top \widetilde{\mathbf{S}}_{t-1}(\boldsymbol{\mu})\boldsymbol{w}}}{\sqrt{n-k+1}} t_2 \,,$$

where $t_2 \sim t_1(n-k+1, 0, 1)$ is independent of $\boldsymbol{\mu}$ and $\mathbf{X}_{(t-1)}$.

Finally, from the properties of the multivariate t-distribution, we obtain

$$\boldsymbol{w}^\top \boldsymbol{\mu} - \boldsymbol{w}^\top \overline{\mathbf{x}}_{t-1} \sim t_1 \left(n-k, 0, \frac{\boldsymbol{w}^\top \mathbf{S}_{t-1} \boldsymbol{w}}{n(n-k)} \right) \,,$$

and, consequently,

$$X_{p,t} \stackrel{d}{=} \boldsymbol{w}^\top \overline{\mathbf{x}}_{t-1} + \sqrt{\boldsymbol{w}^\top \mathbf{S}_{t-1} \boldsymbol{w}} \left(\frac{t_1}{\sqrt{n(n-k)}} + \sqrt{1 + \frac{t_1^2}{n-k}} \frac{t_2}{\sqrt{n-k+1}} \right) \,,$$

where t_1 and t_2 are independent with $t_1 \sim t_{n-k}$ and $t_2 \sim t_{n-k+1}$. □

Proof of Corollary 2: In using the stochastic representation given in Theorem 17 and the properties of the t-distribution, we get

$$\begin{aligned}
\mathbb{E}(\mathbf{X}_t | \mathbf{x}_{(t-1)}) &= \boldsymbol{w}^\top \overline{\mathbf{x}}_{t-1} + \sqrt{\boldsymbol{w}^\top \mathbf{S}_{t-1} \boldsymbol{w}} \left(\frac{\mathbb{E}(t_1)}{\sqrt{n(n-k)}} + \mathbb{E} \left(\sqrt{1 + \frac{t_1^2}{n-k}} \right) \frac{\mathbb{E}(t_2)}{\sqrt{n-k+1}} \right) \\
&= \boldsymbol{w}^\top \overline{\mathbf{x}}_{t-1}
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{V}ar(\mathbf{X}_t|\mathbf{x}_{(t-1)}) = \boldsymbol{w}^\top \mathbf{S}_{t-1}\boldsymbol{w}\mathbb{V}ar\left(\frac{t_1}{\sqrt{n(n-k)}} + \sqrt{1+\frac{t_1^2}{n-k}}\frac{t_2}{\sqrt{n-k+1}}\right)\\
= \;& \boldsymbol{w}^\top \mathbf{S}_{t-1}\boldsymbol{w}\left(\mathbb{E}\left(\frac{t_1^2}{n(n-k)}\right) + \mathbb{E}\left(\left(1+\frac{t_1^2}{n-k}\right)\frac{t_2^2}{n-k+1}\right)\right.\\
+ \;& \left. 2\mathbb{E}\left(\frac{t_1}{\sqrt{n(n-k)}}\sqrt{1+\frac{t_1^2}{n-k}}\frac{t_2}{\sqrt{n-k+1}}\right)\right)\\
= \;& \boldsymbol{w}^\top \mathbf{S}_{t-1}\boldsymbol{w}\left(\frac{1}{n(n-k)}\mathbb{V}ar(t_1) + \left(1+\frac{1}{n-k}\mathbb{V}ar(t_1)\right)\frac{1}{n-k+1}\mathbb{V}ar(t_2)\right)\\
= \;& \boldsymbol{w}^\top \mathbf{S}_{t-1}\boldsymbol{w}\left(\frac{1}{n(n-k)}\frac{n-k}{n-k-2} + \left(1+\frac{1}{n-k}\frac{n-k}{n-k-2}\right)\frac{1}{n-k+1}\frac{n-k+1}{n-k-1}\right)\\
= \;& \left(\frac{1}{n-k-1} + \frac{2n-k-1}{n(n-k-1)(n-k-2)}\right)\boldsymbol{w}^\top \mathbf{S}_{t-1}\boldsymbol{w}\,.
\end{aligned}
$$

□

Chapter 6

Conclusion

This thesis deals with the estimation risk associated with estimating a combination of a product of two multivariate random variables. The main idea is to endow the parameters of interest, the precision matrix and the mean vector, with suitable priors leading to a posterior distribution which allows to derive a stochastic representation of the product of the precision matrix and the mean vector. Hence, the posterior distribution can be described by this stochastic representation. Fortunately, this stochastic representation is also pleasing from a computational point of view since it is described by random variables which can be sampled easily. Some representations can be rewritten using matrix algebra, leading to higher computational efficiency. This method was applied to a variety of portfolio models and using real data, demonstrating the applicability of the stochastic representations. The strengths were demonstrated particularly in deriving a default probability in chapter 2 or a practical optimal solution to the mean-variance approach in chapter 5.

Of course, there are some drawbacks. The most obvious one is that normally distributed observations are needed in order to formulate the likelihood function. This may limit the range of applications. Furthermore, the approach is limited to flexible but few prior distributions since the derivation of the stochastic representation relies on the specific form of the posterior distribution. But this approach makes estimation risk easily accessible, especially since the stochastic representation allows to derive common estimates like the Bayes-estimates for the covariance matrix of the portfolio weights. But most importantly, the need for the stochastic representation as a true distribution of the random variable of interest is emphasized by comparing the true distribution with the asymptotic distribution of the random variable. Since the deviations are rather strong, it might be necessary to access the true distribution of a product of two multivariate random variables. Chapters 4 and 5 particularly demonstrated that estimation risk regarding the random variables considered in this thesis can be considerably vast.

Clearly, the portfolio models considered in this thesis are not the only possible applications of the presented method. For example, research building on chapter 2 and Bodnar et al. (2015b) Theorem 1 is desirable, where the weights for the multi-period portfolio are derived under return predictability and a VAR(1) dependency of the returns and of the predictable variables. Although this is a challenge but would extend a more realistic portfolio model with a notion of estimation uncertainty. But applications of the method of utilizing such stochastic representations are not restricted to portfolio theory since the poduct of the precision matrix and the mean vector are a rather common structure in statistics. Extending the range of applications to discriminant analysis in a similar way as presented in this thesis would also be an interesting point on a research agenda. And such a list can never be complete without stating the obvious in Bayesian statistics: a proper study on the models hyperparameterization.

Bibliography

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18:338–357.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 63(3):1137–1187.

Avramov, D. and Zhou, G. (2010). Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2:25–47.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55:225–264.

Barry, C. (1974). Portfolio analysis under uncertain means, variances, and covariances. *Journal of Finance*, 29:515–522.

Basak, G. K., Jagannathan, R., and Ma, R. (2005). Estimating the risk in sample efficient portfolios. Technical report, Northwestern University, Evanston, IL.

Basak, S. and Chabakauri, G. (2010). Dynamic mean-variance asset allocation. *Review of Financial Studies*, 23:2970–3016.

Bauder, D., Bodnar, R., Bodnar, T., and Schmid, W. (2018a). Bayesian estimation of the efficient frontier. Technical report.

Bauder, D., Bodnar, T., Mazur, S., and Okhrin, Y. (2017a). Bayesian inference for the tangent portfolio. Under revision in International Journal of Theoretical and Applied Finance.

Bauder, D., Bodnar, T., Mazur, S., and Okhrin, Y. (2017b). Bayesian inference of the multi-period optimal portfolio for an exponential utility. Submitted for publication.

Bauder, D., Bodnar, T., Parolya, N., and Schmid, W. (2018b). Mean-variance analysis: optimal portfolio selection under parameter uncertainty. Technical report.

Bawa, V. S., Brown, S. J., and Klein, R. W. (1979). *Estimation Risk and Optimal Portfolio Choice*. North-Holland Pub. Co.

Bernardo, J. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society B*, 41:113–147.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. Wiley.

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The Review of Financial Studies*, 4:315–342.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48:28–43.

Bodnar, O. and Bodnar, T. (2010). On the unbiased estimator of the efficient frontier. *International Journal of Theoretical and Applied Finance*, 13:1065–1073.

Bodnar, T. (2009). An exact test on structural changes in the weights of the global minimum variance portfolio. *Quantitative Finance*, 9:363–370.

Bodnar, T. and Gupta, A. (2009). Construction and inferences of the efficient frontier in elliptical models. *Journal of the Japan Statistical Society*, 39:193–207.

Bodnar, T., Hautsch, N., and Parolya, N. (2017a). Consistent estimation of the high dimensional efficient frontier. Technical report.

Bodnar, T., Mazur, S., and Okhrin, Y. (2017b). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research*, 256:292–307.

Bodnar, T., Mazur, S., Podgórski, K., and Tyrcha, J. (2016). Tangency portfolio weights for small sample and singular covariance matrix: estimation and test theory. Technical report, Lund University.

Bodnar, T. and Okhrin, Y. (2011). On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory. *Scandinavian Jornal of Statistics*, 38:311–331.

Bodnar, T., Parolya, N., and Schmid, W. (2015a). A closed-form solution of the multi-period portfolio choice problem for a quadratic utility function. *Annals of Operations Research*, 229:121–158.

Bodnar, T., Parolya, N., and Schmid, W. (2015b). On the exact solution of the multi-period portfolio choice problem for an exponential utility under return predictability. *European Journal of Operational Research*, 246:528–542.

Bodnar, T., Parolya, N., and Schmid, W. (2017c). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, To appear.

Bodnar, T. and Schmid, W. (2008a). A test for the weights of the global minimum variance portfolio in an elliptical model. *Metrika*, 67(2):127–143.

Bodnar, T. and Schmid, W. (2008b). Estimation of optimal portfolio compositions for gaussian returns. *Statistics & Decisions*, 26:179–201.

Bodnar, T. and Schmid, W. (2009). Econometrical analysis of the sample efficient frontier. *The European Journal of Finance*, 15:317–335.

Bodnar, T. and Schmid, W. (2011). On the exact distribution of the estimated expected utility portfolio weights: Theory and applications. *Statistics* & *Risk Modeling*, 28:319–342.

Bradley, B. O. and Taqqu, M. S. (2003). Financial risk and heavy tails. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 35 – 103. North-Holland, Amsterdam.

Brandt, M. (2009). Portfolio choice problems. In Ait-Sahalia, Y., H. L., editor, *Handbook of Financial Econometrics: Tools and techniques*, pages 269–336. Elsevier.

Brandt, M. (2010). Portfolio choice problems. In Ait-Sahalia, Y. and Hansen, L., editors, *Handbook of Financial Econometrics*, volume 1, pages 269–336. Tools and Techniques, North Hol- land.

Brandt, M. and Santa-Clara (2006). Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61:2187–2217.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45:21–58.

Brown, L.D., C. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16:101–133.

Brown, S. (1976). *Optimal portfolio choice under uncertainty: A Bayesian approach.* PhD thesis, University of Chicago.

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.

Çanakoğlu, E. and Özekici, S. (2009). Portfolio selection in stochastic markets with exponential utility functions. *Annals of Operations Research*, 166:281–297.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman & Hall/CRC,.

Chopra, V. and Ziemba, W. (1993). The effects of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer Texts in Statistics. Springer.

deFinetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *6. Memorie, Classe di Scienze Fisiche, Mathematiche e Naturali. Academia Nazionale dei Lincei*, pages 251–299.

Elton, E. J., G. M. J. (1974). On the optimality of some multiperiod portfolio selection criteria. *Journal of Business*, 47:231–243.

French, K. and Poterba, J. (1991). Investor diversification and international equity markets. *American Economic Review*, 81(2):222–226.

Frost, P. and Savarino, J. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21:293–305.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57:1121–1152.

Givens, G. H. and Hoeting, J. A. (2012). *Computational Statistics.* John Wiley & Sons.

Golosnoy, V. and Okhrin, Y. (2007). Multivariate shrinkage for optimal portfolio weights. *European Journal of Finance*, 13:441–458.

Golosnoy, V. and Okhrin, Y. (2008). Estimation of optimal portfolio compositions for gaussian returns. *Journal of Economic Behavior and Organization*, 67:718–734.

Gupta, A. and Nagar, D. (2000). *Matrix Variate Distributions*. Chapman and Hall/CRC, Boca Raton.

Gupta, A., Varga, T., and Bodnar, T. (2013). *Elliptically Contoured Models in Statistics and Portfolio Theory*. Springer, second edition.

Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2):263–290.

Hewitt, E. and Savage, L. (1955). Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501.

Hodges, S. and Brealey, R. (1978). Portfolio selection in a dynamic and uncertain world. In Lorie, J. and Brealey, R., editors, *Modern Developments in Investment Management*. Dryden Press, Hinsdale, Illinois, 2nd Edition.

Ingersoll, J. E. (1987). *Theory of Financial Decision Making*. Rowman & Littlefield Publishers.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–139.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 453–461.

Jobson, J. (1991). Confidence regions for the mean-variance efficient set: an alternative approach to estimation risk. *Review of Quantitative Finance and Accounting*, 1:235–257.

Jobson, J. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75:544–554.

Jorion, P. (1985). International portfolio diversification with estimation risk. *Journal of Business*, 58:259–278.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Smith, D. R. (2008). The distribution of the sample minimum-variance frontier. *Management Science*, 54:1364–1380.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42:621– 656.

Kandel, S. and Stambaugh, R. (1996). On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance*, 26:385–424.

Klein, R. and Bawa, V. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3:215–231.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press, Cambridge, United Kingdom.

Lai, T. and Xing, H. (2008). *Statistical Models and Methods for Financial Markets*. Springer Texts in Statistics. Springer, New York.

Li, D. and Ng, W.-L. (2000). Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, 10:387–406.

Litterman, R. (2003). *Modern Investment Management: An Equilibrium Approach*. Wiley.

MacKinlay, A. C. and Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies*, 13:883–916.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7:77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley, New York.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms an Random Variables*. Marcel Dekker, New York.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics*, 50:247–257.

Merton, R. C. (1972). An analytic derivation of the efficient portfolio frontier. *Journal of Financial and Quantitative Analysis*, 7:1851–1872.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics*, 8:323–361.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM.

Michaud, R. (1989). The markowitz optimization enigma: Is optimized optimal. *Financial Analysts Journal*, 45:31–42.

Mossin, J. (1968). Optimal multiperiod portfolio policies. *The Journal of Business*, 41:215–229.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Okhrin, Y. and Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics*, 134:235–256.

Pástor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, 55:179–223.

Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56:335–381.

Pennacchi, G. G. (2008). *Theory of Asset Pricing.* Pearson/Addison-Wesley Boston.

Rachev, S. T., Hsu, J. S. J., Bagasheva, B. S., and Fabozzi, F. J. (2008). *Bayesian Methods in Finance.* Wiley, New Jersey.

Ryll-Nardzewski, C. (1957). Colloqium mathematicae. *On stationary sequences of random variables and the de Finetti's equivalence*, 4:149–156.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51:239–246.

Sekerke, M. (2015). *Bayesian Risk Management: A Guide to Model Risk and Sequential Learning in Financial Markets.* Wiley, New Jersey.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, 5:1–33.

Shanken, J. and Zhou, G. (2007). Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics*, 84:40–86.

Siegel, A. and Woodgate, A. (2007). Performance of portfolios optimized with estimation error. *Management Science*, 53:1005–1015.

Stambaugh, R. F. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics*, 45:285–331.

Sun, D. and Berger, J. (2007). Objective bayesian analysis for the multivariate normal model. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics*, volume 8, pages 525–547. Oxford: University Press.

Tu, J. and Zhou, G. (2010). Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 45:959–986.

Wang, Z. (2005). A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, 18:673–705.

Winkler, R. L. (1973). Bayesian models for forecasting future security prices. *Journal of Financial and Quantitative Analysis*, 8:387–405.

Winkler, R. L. and Barry, C. B. (1975). A Bayesian model for portfolio selection and revision. *Journal of Finance*, 30:179–192.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. John Wiley, New York.